BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number
WO 02/30964 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/US01/31931

(22) International Filing Date: 10 October 2001 (10.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/240,143	10 October 2000 (10.10.2000)	US
60/256,409	18 December 2000 (18.12.2000)	US
60/256,410	18 December 2000 (18.12.2000)	US
60/299,452	20 June 2001 (20.06.2001)	US
09/974,366	10 October 2001 (10.10.2001)	US

(71) Applicants and
(72) Inventors: **HICKMAN, Heather** [US/US]; 700 NW 49th Street, Oklahoma City, OK 73118 (US). **HILDEBRAND, William** [US/US]; 900 Northcreek Drive, Edmond, OK 73034 (US).

(74) Agents: **PALMER, John** et al.; Ladas & Parry, 5670 Wilshire Boulevard, Suite 2100, Los Angeles, CA 90036-5679 (US).

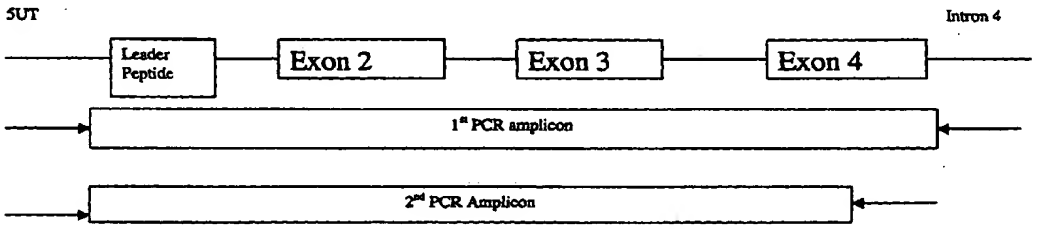
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPARATIVE LIGAND MAPPING FROM MHC POSITIVE CELLS



(57) Abstract: The present invention relates generally to a methodology for the isolation, purification and identification of peptide ligands presented by MHC positive cells. In particular, the methodology of the present invention relates to the isolation, purification and identification of these peptide ligands from soluble class I and class II MHC molecules which may be uninfected, infected, or tumorgenic. The methodology of the present invention broadly allows for these peptide ligands and their comcomittant source proteins thereof to be identified and used as markers for infected versus uninfected cells and/or tumorgenic versus nontumorgenic cells with said identification being useful for marking or targeting a cell for therapeutic treatment or priming the immune response against infected cells.

WO 02/30964 A2

COMPARATIVE LIGAND MAPPING FROM MHC POSITIVE CELLS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of: provisional patent application U.S. Serial No. 60/240,143, filed October 10, 2000, entitled "C-TERMINAL EPITOPE TAGGING FACILITATES COMPARATIVE LIGAND MAPPING FROM MHC CLASS I POSITIVE CELLS"; provisional patent application U.S. Serial No. 60/299,452, filed June 20, 2001, entitled "HIV EPITOPES IDENTIFIED BY THE METHOD OF C-TERMINAL EPITOPE TAGGING FOR COMPARATIVE LIGAND MAPPING FROM MHC CLASS I POSITIVE CELLS"; provisional patent application U.S. Serial No. 60/256,410, filed December 18, 2000, entitled "HLA PRODUCTION FROM GENOMIC DNA"; provisional patent application U.S. Serial No. 60/256,409, entitled "HLA PRODUCTION FROM cDNA" filed December 18, 2000; and provisional patent application U.S. Serial No. not yet assigned, entitled "PRODUCTION OF SOLUBLE HUMAN HLA CLASS I PROTEINS FROM GENOMIC DNA" filed October 9, 2001 all of which are expressly incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This work was funded in part by a contract from the National Institutes of Health: Contract number NO1-AI-95360 entitled "Mapping and Characterization of Viral Epitopes". As such, the Government may own certain rights in and to this application.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to a methodology for the isolation, purification and identification of peptide ligands presented by MHC positive cells. In particular, the methodology of the present invention relates to the isolation, purification and identification of these peptide ligands from soluble class I and class II MHC molecules which may be uninfected, infected, or tumorigenic. The methodology of the present invention broadly allows for these peptide ligands and their concomitant source proteins thereof to be identified and used as markers for infected versus uninfected cells and/or tumorigenic versus nontumorigenic cells with said identification being useful for marking or targeting a cell for therapeutic treatment or priming the immune response against infected cells.

2. Description of the Background Art

Class I major histocompatibility complex (MHC) molecules, designated HLA class I in humans, bind and display peptide antigen ligands upon the cell surface. The peptide antigen ligands presented by the class I MHC molecule are derived from either normal endogenous proteins ("self") or foreign proteins ("nonself") introduced into the cell. Nonself proteins may be products of malignant transformation or intracellular pathogens such as viruses. In this manner, class I MHC molecules convey information regarding the internal fitness of a cell to immune effector cells including but not limited to, $CD8^+$ cytotoxic T lymphocytes (CTLs), which are activated upon interaction with "nonself" peptides, thereby lysing or killing the cell presenting such

"nonself" peptides.

Class II MHC molecules, designated HLA class II in humans, also bind and display peptide antigen ligands upon the cell surface. Unlike class I MHC molecules which are expressed on virtually all nucleated cells, class II MHC molecules are normally confined to specialized cells, such as B lymphocytes, macrophages, dendritic cells, and other antigen presenting cells which take up foreign antigens from the extracellular fluid via an endocytic pathway. The peptides they bind and present are derived from extracellular foreign antigens, such as products of bacteria that multiply outside of cells, wherein such products include protein toxins secreted by the bacteria that often times have deleterious and even lethal effects on the host (e.g. human). In this manner, class II molecules convey information regarding the fitness of the extracellular space in the vicinity of the cell displaying the class II molecule to immune effector cells, including but not limited to, $CD4^+$ helper T cells, thereby helping to eliminate such pathogens the examination of such pathogens is accomplished by both helping B cells make antibodies against microbes, as well as toxins produced by such microbes, and by activating macrophages to destroy ingested microbes.

Class I and class II HLA molecules exhibit extensive polymorphism generated by systematic recombinatorial and point mutation events; as such, hundreds of different HLA types exist throughout the world's population, resulting in a large immunological diversity. Such extensive HLA diversity throughout the population results in tissue or organ transplant rejection between individuals as well as differing susceptibilities and/or resistances to infectious diseases. HLA molecules also contribute significantly to

autoimmunity and cancer. Because HLA molecules mediate most, if not all, adaptive immune responses, large quantities of pure isolated HLA proteins are required in order to effectively study transplantation, autoimmunity disorders, and for vaccine development.

There are several applications in which purified, individual class I and class II MHC proteins are highly useful. Such applications include using MHC-peptide multimers as immunodiagnostic reagents for disease resistance/autoimmunity; assessing the binding of potentially therapeutic peptides; elution of peptides from MHC molecules to identify vaccine candidates; screening transplant patients for preformed MHC specific antibodies; and removal of anti-HLA antibodies from a patient. Since every individual has differing MHC molecules, the testing of numerous individual MHC molecules is a prerequisite for understanding the differences in disease susceptibility between individuals. Therefore, purified MHC molecules representative of the hundreds of different HLA types existing throughout the world's population are highly desirable for unraveling disease susceptibilities and resistances, as well as for designing therapeutics such as vaccines.

Class I HLA molecules alert the immune response to disorders within host cells. Peptides, which are derived from viral- and tumor-specific proteins within the cell, are loaded into the class I molecule's antigen binding groove in the endoplasmic reticulum of the cell and subsequently carried to the cell surface. Once the class I HLA molecule and its loaded peptide ligand are on the cell surface, the class I molecule and its peptide ligand are accessible to cytotoxic T lymphocytes (CTL). CTL survey the peptides presented by the class I molecule and destroy those cells harboring ligands derived from

infectious or neoplastic agents within that cell.

While specific CTL targets have been identified, little is known about the breadth and nature of ligands presented on the surface of a diseased cell. From a basic science perspective, many outstanding questions have permeated through the art regarding peptide exhibition. For instance, it has been demonstrated that a virus can preferentially block expression of HLA class I molecules from a given locus while leaving expression at other loci intact. Similarly, there are numerous reports of cancerous cells that fail to express class I HLA at particular loci. However, there are no data describing how (or if) the three classical HLA class I loci differ in the immunoregulatory ligands they bind. It is therefore unclear how class I molecules from the different loci vary in their interaction with viral- and tumor-derived ligands and the number of peptides each will present.

Discerning virus- and tumor-specific ligands for CTL recognition is an important component of vaccine design. Ligands unique to tumorigenic or infected cells can be tested and incorporated into vaccines designed to evoke a protective CTL response. Several methodologies are currently employed to identify potentially protective peptide ligands. One approach uses T cell lines or clones to screen for biologically active ligands among chromatographic fractions of eluted peptides. (Cox et al., Science, vol 264, 1994, pages 716-719, which is expressly incorporated herein by reference in its entirety) This approach has been employed to identify peptides ligands specific to cancerous cells. A second technique utilizes predictive algorithms to identify peptides capable of binding to a particular class I molecule based upon previously determined motif and/or individual ligand sequences. (De Groot

et al., Emerging Infectious Diseases, (7) 4, 2001, which is expressly incorporated herein by reference in its entirety) Peptides having high predicted probability of binding from a pathogen of interest can then be synthesized and tested for T cell reactivity in precursor, tetramer or ELISpot assays.

However, there has been no readily available source of individual HLA molecules. The quantities of HLA protein available have been small and typically consist of a mixture of different HLA molecules. Production of HLA molecules traditionally involves growth and lysis of cells expressing multiple HLA molecules. Ninety percent of the population is heterozygous at each of the HLA loci; codominant expression results in multiple HLA proteins expressed at each HLA locus. To purify native class I or class II molecules from mammalian cells requires time-consuming and cumbersome purification methods, and since each cell typically expresses multiple surface-bound HLA class I or class II molecules, HLA purification results in a mixture of many different HLA class I or class II molecules. When performing experiments using such a mixture of HLA molecules or performing experiments using a cell having multiple surface-bound HLA molecules, interpretation of results cannot *directly* distinguish between the different HLA molecules, and one cannot be certain that any particular HLA molecule is responsible for a given result. Therefore, a need existed in the art for a method of producing substantial quantities of individual HLA class I or class II molecules so that they can be readily purified and isolated independent of other HLA class I or class II molecules. Such individual HLA molecules, when provided in sufficient quantity and purity, would provide a powerful tool for studying and measuring

immune responses.

Therefore, there exists a need in the art for improved methods of epitope discovery and comparative ligand mapping for class I and class II MHC molecules, including methods of distinguishing an infected/tumor cell from an uninfected/non-tumor cell. The present invention solves this need by coupling the production of soluble HLA molecules with an epitope isolation, discovery, and direct comparison methodology.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. Overview of 2 stage PCR strategy to amplify a truncated version of the human class I MHC.

FIG. 2. Edman sequence analysis of soluble B*1501, B*1501-HIS and B*1501-FLAG. Residue intensity was categorized as either dominant (3.5-fold or more picomolar increase over previous round) or strong (2.5 to 3.5-fold increase over prior round).

FIG. 3. Representative MS ion maps from soluble B*1501, B*1501-HIS and B*1501-FLAG illustrating ion overlap between the molecules. Pooled, acid-eluted peptides were fractionated by RP-HPLC, and the individual fractions were MS scanned.

FIG. 4. Fragmentation pattern generated by MS/MS on an ion selected rrom fraction 11 of B*1501, B*1501-HIS and B*1501-FLAG peptides illustrating a sequence-level overlap between the three molecules.

FIG. 5. Flow chart of the epitope discovery of C-terminal-tagged sHLA molecules. Class I positive transfectants are infected with a pathogen of choice and sHLA preferentially purified utilizing the tag. Subtractive comparison of MS ion maps yields ions present only in infected cell, which are

then MS/MS sequenced to derive class I epitopes.

FIG. 6. MS ion map from soluble B*0702 SupT1 cells uninfected and infected with HIV MN-1. Pooled, acid-eluted peptides were fractionated by RP-HPLC, and fraction #30 was MS scanned.

FIG. 7. MS ion map similar to FIG. 6 but zoomed in on the area from 482-488 amu to more clearly identify all ions in the immediate area.

FIG. 8. Fragmentation pattern generated by tandem mass spectrometry of the peptide ion 484.72 isolated from infected soluble B*0702 SupT1 cells.

FIG. 9. Results of a PubMed BLAST search with the sequence GPRTAALGLL identified in FIG. 8.

FIG. 10. Summary of Results of Entrez-PubMed search for the word "reticulocalbin".

FIG. 11. Results of a peptide-binding algorithm performed using Parker's Prediction using the entire source protein, reticulocalbin, which generates a list of peptides which are bound by the B*0702 HLA allele.

FIG. 12. Results of a peptide-binding algorithm performed using Rammensee's SYPEITHI Prediction using the entire source protein, reticulocalbin, which generates a list of peptides which are bound by the B*0702 HLA allele.

FIG. 13. Results of a predicted proteasomal cleavage of the complete reticulocalbin protein using the cleavage predictor PaProC.

FIG. 14. Results of a predicted proteasomal cleavage of the complete reticulocalbin protein using the cleavage predictor NetChop 2.0.

FIG. 15. Several high affinity peptides deriving from reticulocalbin were identified as peptides predicted to be presented by HLA-A*0201 and A*0101.

FIG. 16. MS ion maps from soluble B*0702 uninfected SupT1 cells of fractions 29 and 31 to determine that ion 484.72 was not present.

FIG. 17. Fragmentation patterns of soluble B*0702 uninfected SupT1 cells fraction 30 ion 484.72 under identical MS collision conditions to illustrate the absence of the reticulocalbin peptide in the uninfected cells.

FIG. 18. Comparison of the MS/MS fragmentation patterns of synthetic peptide GPRTAALGLL and peptide ion 484.72 isolated from infected soluble B*0702 SupT1 cells.

DETAILED DESCRIPTION OF THE INVENTION

Before explaining at least one embodiment of the invention in detail by way of exemplary drawings, experimentation, results, and laboratory procedures, it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of the components set forth in the following description or illustrated in the drawings, experimentation and/or results. The invention is capable of other embodiments or of being practiced or carried out in various ways. As such, the language used herein is intended to be given the broadest possible scope and meaning; and the embodiments are meant to be exemplary - not exhaustive. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

The present invention generally relates to a method of epitope discovery and comparative ligand mapping as well as methods of distinguishing infected/tumor cells from uninfected/non-tumor cells. The present method broadly includes the following steps: (1) providing a cell line

containing a construct that encodes an individual soluble class I or class II MHC molecule (wherein the cell line is capable of naturally processing self or nonself proteins into peptide ligands capable of being loaded into the antigen binding grooves of the class I or class II MHC molecules); (2) culturing the cell line under conditions which allow for expression of the individual soluble class I or class II MHC molecule from the construct, with such conditions also allowing for the endogenous loading of a peptide ligand (from the self or non-self processed protein) into the antigen binding groove of each individual soluble class I or class II MHC molecule prior to secretion of the soluble class I or class II MHC molecules having the peptide ligands bound thereto; and (4) separating the peptide ligands from the individual soluble class I or class II MHC molecules.

The methods of the present invention may, in one embodiment, utilize a method of producing MHC molecules (from genomic DNA or cDNA) that are secreted from mammalian cells in a bioreactor unit. Substantial quantities of individual MHC molecules are obtained by modifying class I or class II MHC molecules so that they are capable of being secreted, isolated, and purified. Secretion of soluble MHC molecules overcomes the disadvantages and defects of the prior art in relation to the quantity and purity of MHC molecules produced. Problems of quantity are overcome because the cells producing the MHC do not need to be detergent lysed or killed in order to obtain the MHC molecule. In this way the cells producing secreted MHC remain alive and therefore continue to produce MHC. Problems of purity are overcome because the only MHC molecule secreted from the cell is the one that has specifically been constructed to be secreted. Thus, transfection of vectors encoding such

secreted MHC molecules into cells which may express endogenous, surface bound MHC provides a method of obtaining a highly concentrated form of the transfected MHC molecule as it is secreted from the cells. Greater purity is assured by transfecting the secreted MHC molecule into MHC deficient cell lines.

Production of the MHC molecules in a hollow fiber bioreactor unit allows cells to be cultured at a density substantially greater than conventional liquid phase tissue culture permits. Dense culturing of cells secreting MHC molecules further amplifies the ability to continuously harvest the transfected MHC molecules. Dense bioreactor cultures of MHC secreting cell lines allow for high concentrations of individual MHC proteins to be obtained. Highly concentrated individual MHC proteins provide an advantage in that most downstream protein purification strategies perform better as the concentration of the protein to be purified increases. Thus, the culturing of MHC secreting cells in bioreactors allows for a continuous production of individual MHC proteins in a concentrated form.

The method of producing MHC molecules utilized in the present invention begins by obtaining genomic or complementary DNA which encodes the desired MHC class I or class II molecule. Alleles at the locus which encode the desired MHC molecule are PCR amplified in a locus specific manner. These locus specific PCR products may include the entire coding region of the MHC molecule or a portion thereof. In one embodiment a nested or hemi-nested PCR is applied to produce a truncated form of the class I or class II gene so that it will be secreted rather than anchored to the cell surface. In another embodiment the PCR will directly truncate the MHC

molecule.

Locus specific PCR products are cloned into a mammalian expression vector and screened with a variety of methods to identify a clone encoding the desired MHC molecule. The cloned MHC molecules are DNA sequenced to insure fidelity of the PCR. Faithful truncated clones of the desired MHC molecule are then transfected into a mammalian cell line. When such cell line is transfected with a vector encoding a recombinant class I molecule, such cell line may either lack endogenous class I MHC molecule expression or express endogenous class I MHC molecules. One of ordinary skill of the art would note the importance, given the present invention, that cells expressing endogenous class I MHC molecules may spontaneously release MHC into solution upon natural cell death. In cases where this small amount of spontaneously released MHC is a concern, the transfected class I MHC molecule can be "tagged" such that it can be specifically purified away from spontaneously released endogenous class I molecules in cells that express class I molecules. For example, a DNA fragment encoding a HIS tail may be attached to the protein by the PCR reaction or may be encoded by the vector into which the PCR fragment is cloned, and such HIS tail, therefore, further aids in the purification of the class I MHC molecules away from endogenous class I molecules. Tags beside a histidine tail have also been demonstrated to work, and one of ordinary skill in the art of tagging proteins for downstream purification would appreciate and know how to tag a MHC molecule in such a manner so as to increase the ease by which the MHC molecule may be purified.

Cloned genomic DNA fragments contain both exons and introns as well

as other non-translated regions at the 5' and 3' termini of the gene. Following transfection into a cell line which transcribes the genomic DNA (gDNA) into RNA, cloned genomic DNA results in a protein product thereby removing introns and splicing the RNA to form messenger RNA (mRNA), which is then translated into an MHC protein. Transfection of MHC molecules encoded by gDNA therefore facilitates reisolation of the gDNA, mRNA/cDNA, and protein. Production of MHC molecules in non-mammalian cell lines such as insect and bacterial cells requires cDNA clones, as these lower cell types do not have the ability to splice introns out of RNA transcribed from a gDNA clone. In these instances the mammalian gDNA transfectants of the present invention provide a valuable source of RNA which can be reverse transcribed to form MHC cDNA. The cDNA can then be cloned, transferred into cells, and then translated into protein. In addition to producing secreted MHC, such gDNA transfectants therefore provide a ready source of mRNA, and therefore cDNA clones, which can then be transfected into non-mammalian cells for production of MHC. Thus, the present invention which starts with MHC genomic DNA clones allows for the production of MHC in cells from various species.

A key advantage of starting from gDNA is that viable cells containing the MHC molecule of interest are not needed. Since all individuals in the population have a different MHC repertoire, one would need to search more than 500,000 individuals to find someone with the same MHC complement as a desired individual – such a practical example of this principle is observed when trying to find a donor to match a recipient for bone marrow transplantation. Thus, if it is desired to produce a particular MHC molecule

for use in an experiment or diagnostic, a person or cell expressing the MHC allele of interest would first need to be identified. Alternatively, in the method of the present invention, only a saliva sample, a hair root, an old freezer sample, or less than a milliliter (0.2 ml) of blood would be required to isolate the gDNA. Then, starting from gDNA, the MHC molecule of interest could be obtained via a gDNA clone as described herein, and following transfection of such clone into mammalian cells, the desired protein could be produced directly in mammalian cells or from cDNA in several species of cells using the methods of the present invention described herein.

Current experiments to obtain an MHC allele for protein expression typically start from mRNA, which requires a fresh sample of mammalian cells that *express* the MHC molecule of interest. Working from gDNA does not require gene expression or a fresh biological sample. It is also important to note that RNA is inherently unstable and is not as easily obtained as is gDNA. Therefore, if production of a particular MHC molecule starting from a cDNA clone is desired, a person or cell line that is expressing the allele of interest must traditionally first be identified in order to obtain RNA. Then a fresh sample of blood or cells must be obtained; experiments using the methodology of the present invention show that $\geq$5 milliliters of blood that is less than 3 days old is required to obtain sufficient RNA for MHC cDNA synthesis. Thus, by starting with gDNA, the breadth of MHC molecules that can be readily produced is expanded. This is a key factor in a system as polymorphic as the MHC system; hundreds of MHC molecules exist, and not all MHC molecules are readily available. This is especially true of MHC molecules unique to isolated populations or of MHC molecules unique to

ethnic minorities. Starting class I or class II MHC molecule expression from the point of genomic DNA simplifies the isolation of the gene of interest and insures a more equitable means of producing MHC molecules for study; otherwise, one would be left to determine whose MHC molecules are chosen and not chosen for study, as well as to determine which ethnic population from which fresh samples cannot be obtained and therefore should not have their MHC molecules included in a diagnostic assay.

While cDNA may be substituted for genomic DNA as the starting material, production of cDNA for each of the desired HLA class I types will require hundreds of different, HLA typed, viable cell lines, each expressing a different HLA class I type. Alternatively, fresh samples are required from individuals with the various desired MHC types. The use of genomic DNA as the starting material allows for the production of clones for many HLA molecules from a single genomic DNA sequence, as the amplification process can be manipulated to mimic recombinatorial and gene conversion events. Several mutagenesis strategies exist whereby a given class I gDNA clone could be modified at either the level of gDNA or at the cDNA resulting from this gDNA clone. The process of producing MHC molecules utilized in the present invention does not require viable cells, and therefore the degradation which plagues RNA is not a problem.

The soluble class I MHC proteins produced by the method described herein is utilized in the methods of epitope discovery and comparative ligand mapping of the present invention. The methods of epitope discovery and comparative ligand mapping described herein which utilize secreted individual MHC molecules have several advantages over the prior art, which utilized

MHC from cells expressing multiple membrane-bound MHCs. While the prior art method could distinguish if an epitope was presented on the surface of a cell, this prior art method is unable to directly distinguish in which specific MHC molecule the peptide epitope was bound. Lengthy purification processes might be used to try and obtain a single MHC molecule, but doing so limits the quantity and usefulness of the protein obtained. The novelty and flexibility of the current invention is that individual MHC specificities can be utilized in sufficient quantity through the use of recombinant, soluble MHC proteins.

Class I and class II MHC molecules are really a trimolecular complex consisting of an alpha chain, a beta chain, and the alpha/beta chain's peptide cargo (i.e. peptide ligand) which is presented on the cell surface to immune effector cells. Since it is the peptide cargo, and not the MHC alpha and beta chains, which marks a cell as infected, tumorigenic, or diseased, there is a great need to identify and characterize the peptide ligands bound by particular MHC molecules. For example, characterization of such peptide ligands greatly aids in determining how the peptides presented by a person with MHC-associated diabetes differ from the peptides presented by the MHC molecules associated with resistance to diabetes. As stated above, having a sufficient supply of an individual MHC molecule, and therefore that MHC molecule's bound peptides, provides a means for studying such diseases. Because the method of the present invention provides quantities of MHC protein previously unobtainable, unparalleled studies of MHC molecules and their important peptide cargo can now be facilitated.

Therefore, the present invention is also related to methods of epitope

discovery and comparative ligand mapping which can be utilized to distinguish infected/tumor cells from uninfected/non-tumor cells by unique epitopes presented by MHC molecules in the disease or non-disease state.

Creation of sHLA molecules from genomic DNA (gDNA)

1. Genomic DNA Extraction. 200 µl of sample either blood, plasma, serum, buffy coat, body fluid or up to $5x10^6$ lymphocytes in 200 µl Phosphate buffered saline were used to extract genomic DNA using the QIAamp® DNA Blood Mini Kit blood and body fluid spin protocol. Genomic DNA quality and quantity was assessed using optical density readings at 260nm and 280nm.

2.1 PCR Strategy. Primers were designed for HLA-A, -B and –C loci in order to amplify a truncated version of the human class I MHC using a 2 stage PCR strategy. The first stage PCR uses a primer set that amplify from the 5′ Untranslated region to Intron 4. This amplicon is used as a template for the second PCR which results in a truncated version of the MHC Class I gene by utilizing a 3′ primer that sits down in exon 4, the 5′ primer remains the same as the 1st PCR. An overview can be seen in FIG. 1. The primers for each locus are listed in TABLE I. Different HLA-B locus alleles require primers with different restriction cut sites depending on the nucleotide sequence of the allele. Hence there are two 5′ and two 3′ truncating primers for the –B locus.

TABLE I

Primer name	Sequence 5'-3'	Locus	Cut site	Annealing site	Seq. ID NO.
PP5UTA	GCGCTCTAGACCCAGACGCCGAGGATGGCC	A	XbaI	5UT	1
3PPI4A	GCCCTGACCCTGCTAAAGGT	A		Intron 4	2
PP5UTB	GCGCTCTAGACCACCCGGACTCAGAATCTCCT	B	XbaI	5UT	3
3PPI4B	TGCTTTCCCTGAGAAGAGAT	B		Intron 4	4
5UTB39	AGGCGAATTCCAGAGTCTCCTCAGACGCG	B*39	EcoRI	5UT B39	5
5PKCE	GGGCGAATTCCCGCCGCCACCATGCGGGTCATGGCGCC	C	EcoRI	5UT	6
3PPI4C	TTCTGCTTTCCTGAGAAGAC	C		Intron 4	7
PP5UT	GGGCGAATTCGGACTCAGAATCTCCCCAGACGCCGAG	B	EcoRI	5UT	8
PP3PEI	CCGCGAATTCTCATCTCAGGGTGAGGGGCT	A,B,C	EcoRI	Exon 4	9
PP3PEIH	CCGCAAGCTTTCATCTCAGGGTGAGGGGCT	A,B,C	HindIII	Exon 4	10
3PEIHC7	CCGCAAGCTTTCAGCTCAGGGTGAGGGGCT	Cw*07	HindIII	Exon 4	11

2.2 Primary PCR. Materials: An Eppendorf Gradient Mastercycler is used for all PCR. (1) H_2O:Dionized ultrafiltered water (DIUF) Fisher Scientific, W2-4,41. (2) PCR nucleotide mix (10 m*M* each deoxyribonucleoside triphosphate [dNTP]), Boehringer Manheim, #1814, 362. (3) 10X *Pfx* Amplification buffer, pH 9.0, GibcoBRL®, part # 52806, formulation is proprietary information. (4) 50mM $MgSO_4$, GibcoBRL®, part #52044. (5) Platinum® *Pfx* DNA Polymerase (B Locus only), GibcoBRL®, 11708-013. (6) *Pfu* DNA Polymerase (A and C Locus), Promega, M7741. (7) *Pfu* DNA Polymerase 10x reaction Buffer with $MgSO_4$, 200m*M* Tris-HCL,pH 8.8,100mM KCl, 100m*M* $(NH_4)_2SO_4$, 20m*M* $MgSO_4$, 1mg/ml nuclease free BSA,1% Triton®X-100. (8) Amplification primers (in ng/µl) (see TABLE I): A locus: 5' sense PP5UTA (300); 3'antisense PPI4A (300); B locus (Not B*39's): sense PP5UTB (300); antisense PPI4B (300); B locus (B*39's): sense 5UTB39 (300); antisense PPI4B (300); C Locus : sense 5PKCE (300); antisense PPI4C (300). (9) gDNA Template.

2.3 Secondary PCR (also used for colony PCR). (1) H_2O:Dionized ultrafiltered water (DIUF) Fisher Scientific, W2-4,41. (2) PCR nucleotide mix (10 m*M* each deoxyribonucleoside triphosphate [dNTP]), Boehringer Manheim, #1814, 362. (3) *Pfu* DNA Polymerase (A and C Locus), Promega, M7741. (4) *Pfu* DNA Polymerase 10x reaction Buffer with $MgSO_4$, 200m*M* Tris-HCL,pH 8.8,100mM KCl, 100m*M* $(NH_4)_2SO_4$, 20m*M* $MgSO_4$, 1mg/ml nuclease free BSA,1% Triton®X-100. (5) Amplification primers (in ng/µl) see TABLE I: A-locus: 5' sense PP5UTA (300), 3' antisense PP3PEI (300); B-locus: sense PP5UTB (300), antisense PP3PEI (300); B-locus: sense PP5UT (300), antisense

PP3PEIH (300); B-locus B39's: sense 5UTB39 (300), antisense PP3PEIH (300); C-locus: sense 5PKCE (300), antisense PP3PEI (300); C-locus Cw*7's: sense 5PKCE (300), antisense 3PEIHC7 (300). (6) Template 1:100 dilution of the primary PCR product.

2.4 Gel Purification of PCR products and vectors. (1) Dark Reader Tansilluminator Model DR-45M, Clare Chemical Research. (2) SYBR Green, Molecular Probes Inc. (3) Quantum Prep Freeze 'N Squeeze DNA Gel Rxtraction Spin Columns, Bio-Rad Laboratories, 732-6165.

2.5 Restriction digests, Ligation and Transformation. (1) Restriction enzymes from New England Biolabs: (a) EcoR I #R0101S; (b) *Hind* III #R0104S; (c) Xba I #R0145S. (2) T4 DNA Ligase, New England Biolabs, #M0202S. (3) pcDNA3.1(-); Invitrogen Corporation, V795-20. (4) 10x Buffers from New England Biolabs: (a) *EcoR* I buffer, 500mM NaCl, 1000mM Tris-HCL, 10mM $MgCL_2$, 0.25% Triton-X 100, pH 7.5; (b) T4 DNA ligase buffer, 500mM Tris-HCL,100mM $MgCL_2$, 100mM DTT, 10mM ATP, 250ug/ml BSA, pH 7.5; (c) NEB buffer 2, 500mM NaCl, 100 mM Tris-HCl, 100mM $MgCl_2$, 10mM DDT, pH 7.9. (5) 100x BSA, New England Biolabs. (6) Z-Competent *E.* coli Transformation Buffer Set, Zymo Research, T3002. (7) *E.* coli strain JM109. (8) LB Plates with 100 µg/ml ampicillin. (9) LB media with 100 µg/ml ampicillin

2.6 Plasmid Extraction. Wizard Plus SV minipreps, Promega, #A1460.

2.7 Sequencing of Clones. (1) Thermo Sequenase Primer Cycle Sequencing Kit, Amersham Pharmacia Biotech, 25-2538-01. (2) CY5 labelled primers

(see TABLE II). (3) AlfExpress automated DNA sequencer, Amersham Pharmacia Biotech.

TABLE II

Primer Name	Sequence 5'-3'	Seq. ID NO:
T7Prom	TAATACGACTCACTATAGGG	12
BGHrev	TAGAAGGCACAGTCGAGG	13
PPI2E2R	GTCGTGACCTGCGCCCC	14
PPI2E2F	TTTCATTTTCAGTTTAGGCCA	15
ABCI3E4F	GGTGTCCTGTCCATTCTCA	16

2.8 Gel Casting. (1) PagePlus 40% concentrate, Amresco, E562, 500ml. (2) Urea, Amersham Pharmacia Biotech, 17-0889-01,500g. (3) 3 N'N'N'N'-tetramethylethyleneiamine (TEMED), APB. (4) Ammonium persulphate (10% solution), APB. (5) Boric acid,APB. (6) EDTA-disodium salt, APB. (7) Tris, APB. (8) Bind-Saline, APB. (9) Isopropanol, Sigma. (10) Glacial Acetic Acid, Fisher Biotech. (11) DIUF water, Fisher Scientific. (12) EtOH 200-proof.

2.9 Plasmid Preparation for Electroporation. Qiagen Plasmid Midi kit, Qiagen Inc., 12143

3.0 Electroporation. (1) Biorad Gene Pulser with capacitance extender, Bio-Rad Laboratories. (2) Gene Pulser Cuvette, Bio-Rad Laboratories. (3) Cytomix: 120mM KCl, 0.15mM $CaCl_2$, 10mMK_2HPO_4/KH_2PO_4, pH 7.6, 25mM Hepes,pH 7.6, 2mM EGTA, pH 7.6, 5mM $MgCl_2$, pH 7.6 with KOH. (4) RPMI 1640+ 20% Foetal Calf Serum + Pen/strep. (5) Haemacytometer. (6) Light Microscope. (7) CO_2 37° Incubator. (8) Cells in log phase.

Primary PCR

c. A-Locus and C-Locus

10x Pfu buffer	5 µl
5' Primer (300ng/µl)	1 µl
3' Primer (300ng/µl)	1 µl
dNTP's (10mM each)	1 µl
gDNA (50ng/µl)	10 µl
DIUF H_20	31.4 µl
Pfu DNA Polymerase	0.6 µl

96°C 2 min.
95°C 1 min
58°C 1 min } x35
73°C 5 min

d. B-locus

10x Pfx buffer	5 µl
5' Primer (300ng/µl)	1 µl
3' Primer (300ng/µl)	1 µl
dNTP's (10mM each)	1.5 µl
$MgSO_4$ (50mM)	1 µl
gDNA (100ng/µl)	1 µl
DIUF H_2O	40 µl
Pfx DNA Polymerase	0.5 µl

94°C 2 min.
94°C 1 min
60°C 1 min } x35
68°C 3.5 min
68°C 5 min

Gel Purification of PCR (all PCR and plasmids are gel purified)

Mix primary PCR with 5 µl of 1x SYBR green and incubate at room temperature for 15 minutes then load on a 1% agarose gel. Visualize on the Dark Reader and purify using the Quantum Prep Freeze and Squeeze extraction kit according to the manufacturers instructions.

Secondary PCR

A, B and C Loci

10x Pfu buffer	5 µl
5' Primer (300ng/µl)	0.5 µl
3' Primer (300ng/µl)	0.5 µl
dNTP's (10mM each)	1 µl
1:100 1° PCR	10 µl
DIUF H_2O	37.5 µl
Pfu DNA Polymerase	0.5 µl

96°C 2 min.
95°C 1 min
58°C 1 min } x35
73°C 4 min
73°C 7 min

Restriction digests

2° PCR (gel purified)	30 µl

Restriction enzyme 2	X µl
10x buffer	5 µl
100x BSA	0.5 µl
DIUF H_2O	10.5 µl

The enzymes used will be determined by the cut sites incorporated into the PCR primers for each individual PCR. The expression vector pcDNA3.1(-) will be cut in a similar manner.

Ligation

PcDNA3.1(-) cut with same enzymes as PCR	50ng
Cut PCR	100ng
10x T4 DNA ligase buffer	2 µl
T4 DNA Ligase	1 µl
DIUF H_20	up to 20 µl

Transformation

Transform JM109 using competent cells made using Z-competent *E.coli* Transformation Kit and Buffer Set and follow the manufacturers instructions.

Colony PCR

This will check for insert in any transformed cells. Follow the same protocol for the secondary PCR.

Mini Preps of colonies with insert

Use the Wizard Plus SV minipreps and follow the manufacturers instructions. Make glycerol stocks before beginning extraction protocol.

Sequencing of positive clones

Using the Thermo Sequenase Primer Cycle Sequencing Kit

A,C,G or T mix	3 µl
CY5 Primer 1pm/µl	1 µl
DNA template 100ng/µl	5 µl
96oC 2 min	
96oC 30 sec } x25	
61oC 30 sec }	

Add 6 µl formamide loading buffer and load 10 µl onto sequencing gel. Analyse sequence for good clones with no misincorporations.

Midi Preps

Prepare plasmid for electroporation using the Qiagen Plasmid Midi Kit according to the manufacturers instructions.

Electroporation

Electroporations are performed as described in "The Bw4 public epitope of HLA-B molecules confers reactivity with natural killer cell clones that express NKb1, a putative HLA receptor. Gumperz, J.E., V. Litwin, J.H. Phillips, L.L. Lanier and P. Parham. J. Exp. Med. 181:1133-1144, 1995, which is expressly incorporated herein by reference."

Screening for production of Soluble HLA

An ELISA is used to screen for the production of soluble HLA. For biochemical analysis, monomorphic monoclonal antibodies are particularly useful for identification of HLA locus products and their subtypes.

W6/32 is one of the most common monoclonal antibodies (mAb) used to characterize human class I major histocompatibility complex (MHC) molecules. It is directed against monomorphic determinants on HLA-A, -B and -C HCs, which recognizes only mature complexed class I molecules and recognizes a conformational epitope on the intact MHC molecule containing both beta2-microglobulin (β2m) and the heavy chain (HC). W6/32 binds a compact epitope on the class I molecule that includes both residue 3 of beta2m and residue 121 of the heavy chain (Ladasky JJ, Shum BP, Canavez F, Seuanez HN, Parham P. Residue 3 of beta2-microglobulin affects binding of class I MHC molecules by the W6/32 antibody. Immunogenetics 1999

Apr;49(4):312-20.). The constant portion of the molecule W6/32 binds to is recognized by CTLs and thus can inhibit cytotoxicity. The reactivity of W6/32 is sensitive to the amino terminus of human beta2-microglobulin (Shields MJ, Ribaudo RK. Mapping of the monoclonal antibody W6/32: sensitivity to the amino terminus of beta2-microglobulin. Tissue Antigens 1998 May;51(5):567-70). HLA-C could not be clearly identified in immunoprecipitations with W6/32 suggesting that HLA-C locus products may be associated only weakly with b2m, explaining some of the difficulties encountered in biochemical studies of HLA-C antigens [Stam, 1986 #1]. The polypeptides correlating with the C-locus products are recognized far better by HC-10 than by W6/32 which confirms that at least some of the C products may be associated with b2m more weakly than HLA-A and -B. W6/32 is available biotinylated (Serotec MCA81B) offering additional variations in ELISA procedures.

HC-10 is reactive with almost all HLA-B locus free heavy chains. The A2 heavy chains are only very weakly recognized by HC-10. Moreover, HC-10 reacts only with a few HLA-A locus heavy chains. In addition, HC-10 seems to react well with free heavy chains of HLA-C types. No evidence for reactivity of HC-10 with heavy-chain/b2m complex has been obtained. None of the immunoprecipitates obtained with HC-10 contained b2m [Stam, 1986 #1]. This indicates that HC-10 is directed against a site of the HLA class I heavy chain that includes the portion involved in interaction with the β2m. The pattern of HC-10 precipitated material is qualitatively different from that isolated with W6/32.

TP25.99 detects a determinant in the alpha3 domain of HLA-ABC. It is found on denatured HLA-B (in Western) as well as partially or fully folded

HLA-A,B,& C. It doesn't require a peptide or β2m, i.e. it works with the alpha 3 domain which folds without peptide. This makes it useful for HC determination.

Anti-human β2m (HRP) (DAKO P0174) recognizes denatured as well as complexed β2m. Although in principle anti-β2m reagents could be used for the purpose of identification of HLA molecules, they are less suitable when association of heavy chain and β2m is weak. The patterns of class I molecules precipitated with W6/32 and anti-β2m are usually indistinguishable [Vasilov, 1983 #10].

Rabbit anti-β2-microglobulin dissociates β2-microglobulin from heavy chain as a consequence of binding (Rogers, M.J., Appella, E., Pierotti, M. A., Invernizzi, G., and Parmiani, G. (1979) Proc Natl. Acad. Sci. U.S.A. 76, 1415-1419). It also has been reported that rabbit anti-human β2-microglobulin dissociactes β2-microglobulin from HLA heavy chains upon binding (Nakamuro, K., Tanigaki, N., and Pressman, D. (1977) Immunology 32, 139-146.). This anti-human β2m antibody is also available unconjugated (DAKO A0072).

The W6/32-HLA sandwich ELISA. Sandwich assays can be used to study a number of aspects of protein complexes. If antibodies are available to different components of a heteropolymer, a two-antibody assay can be designed to test for the presence of the complex. Using a variation of these assays, monoclonal antibodies can be used to test whether a given antigen is multimeric. If the same monoclonal antibody is used for both the solid phase and the label, monomeric antigens cannot be detected. Such combinations, however, may detect multimeric forms of the antigen. In these

assays negative results may be generated both by multimeric antigen held in unfavorable steric positions as well as by monomeric antigens.

The W6/32 - anti-β2m antibody sandwich assay is one of the best techniques for determining the presence and quantity of sHLA. Two antibody sandwich assays are quick and accurate, and if a source of pure antigen is available, the assay can be used to determine the absolute amounts of antigen in unknown samples. The assay requires two antibodies that bind to non-overlapping epitopes on the antigen. This assay is particularly useful to study a number of aspects of protein complexes.

To detect the antigen (sHLA), the wells of microtiter plates are coated with the specific (capture) antibody W6/32 followed by the incubation with test solutions containing antigen. Unbound antigen is washed out and a different antigen-specific antibody (anti-β2m) conjugated to HRP is added, followed by another incubation. Unbound conjugate is washed out and substrate is added. After another incubation, the degree of substrate hydrolysis is measured. The amount of substrate hydrolyzed is proportional to the amount of antigen in the test solution.

The major advantages of this technique are that the antigen does not need to be purified prior to use and that the assays are very specific. The sensitivity of the assay depends on 4 factors: (1) The number of capture antibody; (2) The avidity of the capture antibody for the antigen; (3) The avidity of the second antibody for the antigen; (4) The specific activity of the labeled second antibody.

Using an ELISA protocol template and label a clear 96-well polystyrene assay plate. Polystyrene is normally used as a microtiter plate. (Because it

is not translucent, enzyme assays that will be quantitated by a plate reader should be performed in polystyrene and not PVC plates).

Coating of the W6/32 is performed in Tris buffered saline (TBS); pH 8.5. A coating solution of 8.0 µg/ml of specific W6/32 antibody in TBS (pH 8.5) is prepared. (blue tube preparation stored at -20°C with a concentration of 0.2 mg/ml and a volume of 1 ml giving 0.2 mg per tube).

TABLE III

No. of plates	Total Volume	W6/32 antibody	TBS pH 8.5
1	10 ml	400 µl	9.6 ml
2	20 ml	800 µl	19.2 ml
3	30 ml	1200 µl	28.8 ml
4	40 ml	1600 µl	38.4 ml
5	50 ml	2000 µl	48.0 ml

Although this is well above the capacity of a microtiter plate, the binding will occur more rapidly. Higher concentrations will speed the binding of antigen to the polystyrene but the capacity of the plastic is only about 100 ng/well (300 ng/cm_2), so the extra protein will not bind. (If using W6/32 of unknown composition or concentration, first titrate the amount of standard antibody solution needed to coat the plate versus a fixed, high concentration of labeled antigen. Plot the values and select the lowest level that will yield a strong signal. Do not include sodium azide in any solutions when horseradish peroxidase is used for detection.

Immediately coat the microtiter plate with 100 µl of antigen solution per well using a multichannel pipet. Standard polystyrene will bind antibodies or antigens when the proteins are simply incubated with the plastic. The

bonds that hold the proteins are non-covalent, but the exact types of interactions are not known. Shake the plate to ensure that the antigen solution is evenly distributed over the bottom of each well. Seal the plate with plate sealers (sealplate adhesive sealing film, nonsterile, 100 per unit; Phenix; LMT-Seal-EX) or sealing tape to Nunc-Immuno™ Modules (# 236366). Incubate at 4°C overnight. Avoid detergents and extraneous proteins. Next day, remove the contents of the well by flicking the liquid into the sink or a suitable waste container. Remove last traces of solution by inverting the plate and blotting it against clean paper toweling. Complete removal of liquid at each step is essential for good performance.

Wash the plate 10 times with Wash Buffer (PBS containing 0.05 % Tween-20) using a multi-channel ELISA washer. After the last wash, remove any remaining Wash Buffer by inverting the plate and blotting it against clean paper toweling. After the W6/32 is bound, the remaining sites on the plate must be saturated by incubating with blocking buffer made of 3% BSA in PBS. Fill the wells with 200 µl blocking buffer. Cover the plates with an adhesive strip and incubate overnight at 4°C. Alternatively, incubate for at least 2 hours at room temperature which is, however, not the standard procedure. Blocked plates may be stored for at least 5 days at 4°C. Good pipetting practice is most important to produce reliable quantitative results. The tips are just as important a part of the system as the pipette itself. If they are of inferior quality or do not fit exactly, even the best pipette cannot produce satisfactory results. The pipette working position is always vertical: Otherwise causing too much liquid to be drawn in. The immersion depth should be only a few millimeters. Allow the pipetting button to retract

gradually, observing the filling operation. There should be no turbulence developed in the tip, otherwise there is a risk of aerosols being formed and gases coming out of solution.

When maximum levels of accuracy are stipulated, prewetting should be used at all times. To do this, the required set volume is first drawn in one or two times using the same tip and then returned. Prewetting is absolutely necessary on the more difficult liquids such as 3% BSA. Do not prewet, if your intention is to mix your pipetted sample thoroughly with an already present solution. However, prewet only for volumes greater than 10 µl. In the case of pipettes for volumes less than 10 µl the residual liquid film is as a rule taken into account when designing and adjusting the instrument. The tips must be changed between each individual sample. With volumes <10 µl special attention must also be paid to drawing in the liquid slowly, otherwise the sample will be significantly warmed up by the frictional heat generated. Then slowly withdraw the tip from the liquid, if necessary wiping off any drops clinging to the outside.

To dispense the set volume hold the tip at a slight angle, press it down uniformly as far as the first stop. In order to reduce the effects of surface tension, the tip should be in contact with the side of the container when the liquid is dispensed. After liquid has been discharged with the metering stroke, a short pause is made to enable the liquid running down the inside of the tip to collect at its lower end. Then press it down swiftly to the second stop, in order to blow out the tip with the extended stroke with which the residual liquid can be blown out. In cases that are not problematic (e.g. aqueous solutions) this brings about a rapid and virtually complete discharge

of the set volume. In more difficult cases, a slower discharge and a longer pause before actuating the extended stroke can help. To determine the absolute amount of antigen (sHLA), sample values are compared with those obtained using known amounts of pure unlabeled antigen in a standard curve.

For accurate quantitation, all samples have to be run in triplicate, and the standard antigen-dilution series should be included on each plate. Pipetting should be preformed without delay to minimize differences in time of incubation between samples. All dilutions should be done in blocking buffer. Thus, prepare a standard antigen-dilution series by successive dilutions of the homologous antigen stock in 3% BSA in PBS blocking buffer. In order to measure the amount of antigen in a test sample, the standard antigen-dilution series needs to span most of the dynamic range of binding. This range spans from 5 to 100 ng sHLA/ml. A stock solution Ê of 1 µg/ml should be prepared, aliquoted in volumes of 300 µl and stored at 4°C. Prepare a 50 ml batch of standard at the time. (New batches need to be compared to the old batch before used in quantitation).

Use a tube of the standard stock solution Ê to prepare successive dilutions. While standard curves are necessary to accurately measure the amount of antigen in test samples, they are unnecessary for qualitative "yes/no" answers. For accurate quantitation, the test solutions containing sHLA should be assayed over a number of at least 4 dilutions to assure to be within the range of the standard curve. Prepare serial dilutions of each antigen test solution in blocking buffer (3% BSA in PBS). After mixing, prepare all dilutions in disposable U-bottom 96 well microtiter plates before adding them to the W6/32-coated plates with a multipipette. Add 150 µl in

each well. To further proceed, remove any remaining blocking buffer and wash the plate as described above. The plates are now ready for sample addition. Add 100 µl of the sHLA containing test solutions and the standard antigen dilutions to the antibody-coated wells.

Cover the plates with an adhesive strip and incubate for exactly 1 hour at room temperature. After incubation, remove the unbound antigen by washing the plate 10x with Wash Buffer (PBS containing 0.05 % Tween-20) as described. Prepare the appropriate developing reagent to detect sHLA. Use the second specific antibody, anti-human β2m-HRP (DAKO P0174 / 0.4 mg/ml) conjugated to Horseradish Peroxidase (HRP). Dilute the anti-human β2m-HRP in a ratio of 1:1000 in 3% BSA in PBS. (Do not include sodium azide in solutions when horseradish peroxidase is used for detection).

TABLE IV

No. of plates	Total Volume	anti-β2m-HRP antibody	3%BSA in PBS
1	10 ml	10 µl	10 ml
2	20 ml	20 µl	2ml
3	30 ml	30 µl	30 ml
4	40 ml	40 µl	40 ml
5	50 ml	50 µl	50 ml

Add 100 µl of the secondary antibody dilution to each well. All dilutions should be done in blocking buffer. Cover with a new adhesive strip and incubate for 20 minutes at room temperature. Prepare the appropriate amount of substrate prior to the wash step. Bring the substrate to room temperature.

OPD (o-Phenylenediamine) is a peroxidase substrate suitable for use in ELISA procedures. The substrate produces a soluble end product that is yellow in color. The OPD reaction is stopped with 3 N H_2SO_4, producing an orange-brown product and read at 492 nm. Prepare OPD fresh from tablets (Sigma, P6787; 2 mg/tablet). The solid tablets are convenient to use when small quantities of the substrate are required. After second antibody incubation, remove the unbound secondary reagent by washing the plate 10x with Wash Buffer (PBS containing 0.05% Tween-20). After the final wash, add 100 µl of the OPD substrate solution to each well and allow to develop at room temperature for 10 minutes. Reagents of the developing system are light-sensitive, thus, avoid placing the plate in direct light. Prepare the 3 N H_2SO_4 stop solution. After 10 minutes, add 100 µl of stop solution per 100 µl of reaction mixture to each well. Gently tap the plate to ensure thorough mixing.

Read the ELISA plate at a wavelength of 490 nm within a time period of 15 minutes after stopping the reaction. The background should be around 0.1. If your background is higher, you may have contaminated the substrate with a peroxidase. If the substrate background is low and the background in your assay is high, this may be due to insufficient blocking. Finally analyze your readings. Prepare a standard curve constructed from the data produced by serial dilutions of the standard antigen. To determine the absolute amount of antigen, compare these values with those obtained from the standard curve.

Creation of Transfectants and Production of Soluble Class I Molecules

Transfectants were established as previously described (Prilliman, KR et al., Immunogenetics 45:379, 1997, which is expressly incorporated herein by reference) with the following modifications: a cDNA clone of B*1501 containing the entire coding region of the molecule was PCR amplified in order to generate a construct devoid of the cytoplasmic domain using primers 5PXI (59-GGGCTCTAGAGGACTCAGAATCTCCCCAGAC GCCGAG-39) and 3PEI (59-CCGCGAATTCTCATCTCAGGGTGAG-39) as shown in TABLE V. Constructs were also created containing a C-terminal epitope tag consisting of either 6 consecutive histidines or the FLAG epitope (Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys). TABLE V Primers utilized to create B*1501-HIS and B*1501-FLAG were 5PXI and 3PEIHIS (59-CCGCGAATTCTCAGTGGTGGTGGTGGTGGTGCCATCTCAGGGTGAG-39) or 3PEIFLAG (59-CCGCGAATTCTCACTTGTCATCGTCGTCCTTGTAATCCCATCTCAGGGTGAG-39). PCR amplicons were purified using a Qiagen Spin PCR purification kit (Qiagen, Levsden, The Netherlands) and cloned into the mammalian expression vector pCDNA 3.1 (Invitrogen, Carlsbad, CA, USA). TABLE V. After confirmation of insert integrity by bidirectional DNA sequencing, constructs were electroporated into the class I negative B-lymphoblastoid cell line 721.221 (Prilliman, KR et al., 1997, previously incorporated herein by reference). Transfectants were maintained in medium containing G418 post-electroporation and subcloned in order to isolate efficient producers of soluble class I as determined by ELISA (Prilliman, KR et al, 1997, previously incorporated herein by reference).

TABLE V

ame	Sequence	Full-length or Truncating	Notes		Seq. ID NO:
	GGGCGTCGACGGACTCAGAA TCTCCCCAGACGCCGAG	either	5' primer, Sal I cut site		17
	GCGCGTCGACCCCAGACGC CGAGGATGGCC	either	5' primer, Sal I cut site	A-locus specific	18
	GGGCTCTAGAGGACTCAGAA TCTCCCCAGACGCCGAG	either	5' primer, Xba I cut site		19
	CCGCGTCGACTCAGATTCTC CCCAGACGCCGAGATG	full-length	5' primer, Sal I cut site	C-locus specific	20
	CCGCAAGCTTAGAAACAAG TCAGGGTT	full-length	3' primer, HindIII cut site	A-locus specific	21
	CCGCAAGCTTGGCAGCTGTC TCAGGCTTTACAAG(CT)G	full-length	3' primer, HindIII cut site	C-locus specific	22
	CCGCAAGCTTTTGGGAGGG AGCACAGGTCAGCGTGGGAAG	full-length	3'primer, HindIII cut site	A-locus specific	23
	CCGCAAGCTTCTGGGGAGGA AACATAGGTCAGCATGGGAAC	full-length	3' primer, HindIII cut site	B-locus specific	24
	CCGCGAATTCTCATCTCAGG GTGAG	truncating	3' primer, EcoRI cut site		25
	CCGCGAATTCTCAGTGGTGG TGGTGGTGGTGCCATCTCAG GGTGAG	truncating	3' primer, EcoRI cut site	adds hexa-histidine tail	26
	CCGCGATTCTCACTTGTCATC GTCGTCCTTGTAATCCCATCT CAGGGTGAG	truncating	3' primer, EcoRI cut site	adds FLAG-epitope	27
	GGGCTCTAGACCGCCGCCAC CATGCGGGTCATGGCGCC	either	5' primer, Xba I cut site	C-locus specific	28

Soluble B*1501, B*1501-HIS, and B*1501-FLAG were produced by culturing established transfectants in CP3000 hollow-fiber bioreactors as previously described by Prilliman *et al*, 1997, which has previously been incorporated herein by reference. Supernatants containing soluble class I molecules were collected in bioreactor harvests and purified on W6/32 affinity columns. At least 2 column purifications were performed per molecule.

Ligand Purification, Edman Sequencing, and Reverse-Phase HPLC Separation of Peptides

Peptide ligands were purified from class I molecules by acid elution (Prilliman, KR et al., Immunogenetics 48:89, 1998 which is expressly incorporated herein by reference) and further separated from heavy and light chains by passage through a stirred cell (Millipore, Bedford, MA, USA) equipped with a 3-Kd cutoff membrane (Millipore). Approximately 1/100 volume of stirred cell flow through containing peptide eluted from either B*1501, B*1501-HIS, or B*1501-FLAG was subjected to 14 cycles of Edman degradation on a 492A pulsed liquid phase protein sequencer (Perkin-Elmer Applied Biosystems Division, Norwalk, CT, USA) without the derivitization of cysteine. Edman motifs were derived by combining from multiple column elutions the picomolar yields of each amino acid and then calculating the fold increase over previous round as described in (Prilliman, KR et al, 1998, previously incorporated herein by reference) and are shown in FIG. 2.

Pooled peptide eluate was separated into fractions by RP-HPLC as previously described (Prilliman, KR et al, 1998, previously incorporated herein by reference). Briefly, 400-mg aliquots of peptides were dissolved in 100 ml of 10% acetic acid and loaded onto a 2.1 3 150 mm C18 column (Michrom

Bioresources, Auburn, CA, USA) using a gradient of 2%–10% acetonitrile with 0.06% TFA for 0.02 min followed by a 10%–60% gradient of the same for 60 min. Fractions were collected automatically at 1-min intervals with a flow rate of 180 ml/min.

Mass Spectrometric Ligand Analysis

RP-HPLC fractions were speed-vacuumed to dryness and reconstituted in 40 ml 50% methanol, 0.5% acetic acid. Approximately 6 ml from selected fractions were sprayed into an API-III triple-quadrupole mass spectrometer (PE Sciex, Foster City, CA, USA) using a NanoES ionization source inlet (Protana, Odense, Denmark). Scans were collected while using the following instrument settings: polarity—positive; needle voltage—1375 V; orifice voltage —65 V; N2 curtain gas—0.6 ml/min; step size— 0.2 amu; dwell time—1.5 ms; and mass range—325–1400. Total ion traces generated from each molecule were compared visually in order to identify ions overlapping between molecules. Following identification of ion matches, individual ions were selected for MS/MS sequencing.

Sequences were predicted using the BioMultiView program (PE Sciex) algorithm predict sequence, and fragmentation patterns further assessed manually. Determinations of ion sequence homology to currently compiled sequences were performed using advanced BLAST searches against the nonredundant, human expressed sequence tag, and unfinished high throughput genomic sequences nucleotide databases currently available through the National Center for Biotechnology Information (National Institutes of Health, Bethesda, MD, USA).

The methodology of the present invention provides a direct comparative analysis of peptide ligands eluted from class I HLA molecules. In order to accomplish such comparative analyses , hollow-fiber bioreactors for class I ligand production were used along with reverse-phase HPLC for fractionating eluted ligands, and mass spectrometry for the mapping and sequencing of peptide ligands. The application of comparative ligand mapping also is applicable to cell lines that express endogenous class I. Prior to peptide sequence determination in class I positive cell lines, the effects of adding a C-terminal epitope tag to transfected class I molecules was found to have no deleterious effects. Either a tag consisting of 6 histidines (6-HIS) or a tag containing the epitope Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (FLAG) was added to the C-terminus of soluble B*1501 through PCR. These constructs were then transfected into class I negative 721.221 cells and peptides purified as previously established (Prilliman, KR et al, 1998, previously incorporated herein by reference). Comparison of the two tailed transfectants with the untailed, soluble B*1501 allowed for the determination that tag addition had no effect on peptide binding specificity of the class I molecule and consequently had no deleterious effects on direct peptide ligand mapping and sequencing.

Edman Motifs

The most common means for discerning ligands presented by a particular class I molecule is Edman sequencing the pool of peptides eluted from that molecule. In order to demonstrate that tailing class I molecules with Cterminal tags does not disrupt endogenous peptide loading, Edman sequences of the peptide pools from B*1501, B*1501-HIS, and B*1501-FLAG

was compared with previously published B*1501 data FIG. 2. Motifs were assigned to each of the various B*1501 molecules as shown in FIG. 2. At the anchor position 2 (P2) a dominant Q and subdominant M was seen in motifs as previously published by Falk *et al.* (Immunogenetics 41:165, 1995) and Barber *et al.* (J Exp Med 184:735, 1996). A more disparate P3 is seen in all molecules with F, K, N, P, R, and Y appearing; these results have also been previously reported by Falk and Barber. Again, a dominant Y and F are seen as the C-terminal anchors at P9 in all three molecules. The motif data for all three molecules are in close accord, therefore, with the published standard motifs.

Mass Spectrometric Profiles

Comparison of motifs for the surface bound, nontailed, and tailed B*1501 molecules identified no substantial differences in the pooled peptides bound by the various forms of B*1501 tested. However, the aim of the present invention is to subtractively compare the individual peptides bound by class I molecules from diseased and healthy cells. Subtractive analysis is accomplished through the comparison of mass spectrometric ion maps and, as such, the ion maps.
of tailed and untailed class I molecules were compared in order to determine the effect of tailing upon comparative peptide mapping.

Peptides derived from tailed and untailed B*1501 were separated into fractions via reverse phase HPLC (RP-HPLC). Each fraction was then scanned using an API-III mass spectrometer in order to identify ions present in each fraction. Overall ion scans from RP-HPLC fractions 9, 10, 11, 18, 19, and 20 were produced and visually compared in order to assess ions representing

peptides overlapping between the three molecules. FIG. 3. depicts a representative section of the ion maps generated from each of the molecules. This comparison shows that the same pattern of ions is produced by the different B*1501 molecules analyzed here. The manual comparison of ion maps from each of the three fractions found little to no difference in the peptides bound by each of the three molecules.

Ligand Sequences

After identification of ion matches in MS chromatograms of each of the three molecules, individual ions were chosen for sequencing by tandem mass spectrometry in order to determine if ions were indeed matched at the peptide-sequence level. Ten ions from each fraction were initially selected for MS/MS sequence generation. Fragmentation patterns for each of the ions from each molecule were manually compared and identical fragmentation patterns were counted as peptide-sequence level matches, as illustrated in FIG. 4. Of the peptide fragmentation patterns examined, 52/57 (91%) were exact matches between the untailed molecules and the 6-HIStailed protein (TABLE VI). A more disparate pattern of fragmentation was identified in the FLAG-tailed ions selected for MS/MS sequencing: of the 57 ions selected for MS/MS fragmentation comparison, 39 (70%) fragmentation patterns matched between the FLAG-tailed and untailed molecules. Overall, 91 out of 113 (81%) spectra examined were in accord between the tailed molecules and soluble B*1501.

TABLE VI

Molecules		Ions Examined	Ion Matches	Percent Matched
B*1501-HIS		57	52	91%
B*1501-FLAG		56	39	70%
B*1501-Tagged		113	91	81%

Several ligand sequences were clearly determined from the fragmentation patterns produced. The ligand QGLISRGYSY, deriving from human periplakin, was sequenced from those peptides eluted in fraction 18. A second ligand, AVRDISEASVF, an 11-mer matching a span of the 40S ribosomal protein S26, was identified in fraction 20. Notably, these two peptides lacked the strong consensus glutamine expected by the motif data, a phenomenon previously reported by our laboratory when sequencing B*1501-eluted ligands (Prilliman, KR et al, 1997, previously incorporated herein by reference). Both these ligands, however, terminate with an aromatic tyrosine or phenylalanine; these amino acids were both predicted to be strong anchors by Edman sequencing data and by previously published observations (Prilliman, KR et al, 1998, previously incorporated herein by reference).

One embodiment of the present invention contemplates characterizing peptide ligands bound by a given class I molecule by transfecting that molecule into a class I negative cell line and affinity purification of the class I molecule and bound peptide. Complications arise, however, when cell lines are chosen for study that already possess class I molecules. In this case,

antibodies specific for one class I molecule must be used to selectively purify that class I molecule from others expressed by the cell. Because allele-specific antibodies recognize epitopes in and around the peptide binding groove, variations in the peptides found in the groove can alter antibody affinity for the class I molecule (Solheim, JC et al., J Immunol 151:5387, 1993; and Bluestone, JA et al., J Exp Med 176:1757, 1992). Altered antibody recognition can, in turn, bias the peptides available for elution and subsequent sequence analysis.

In order to selectively purify from a class I positive cell a transfected class I molecule and its peptide ligands in an unbiased way, it was necessary to alter the embodiment for class I purification in a non-class I positive cell. The C-terminal addition of a FLAG and 6-HIS tag to a class I molecule that had already been extensively characterized, B*1501 was shown to have little or no effect on peptide binding. This methodology was designed to allow purification of a single class I specificity from a complex mixture of endogenously expressed class I molecules. Ligands eluted from the tailed and untailed B*1501 molecule were compared to assess the effect of a tail addition on the peptide repertoire.

Pooled Edman sequencing is the commonly used method to determine the binding fingerprint of a given molecule, and this methodology was used to ascertain the large-scale effect of tail addition upon peptide binding. We subjected 1/100 of the peptides eluted from each class I MHC molecule to Edman degradation and derived motifs for each of the molecules. Both the HIS- and FLAG-tailed motifs matched published motifs for the soluble and membrane-bound B*1501. Each of the molecules exhibited motifs bearing a

dominant P2 anchor of Q, a more disparate P3 in which multiple residues could be found, and another dominant anchor of Y or F at P9. Small differences in the picomolar amounts of each of the amino acids detected during Edman sequencing have been noted previously in consecutive runs with the same molecule and most likely reflect differences in cell handling and/or peptide isolation rather than disparities in bound peptides. Highly similar peptide motifs indicated that the peptide binding capabilities of class I MHC molecules are not drastically altered by the addition of a tag.

In order to insure the ligands were not skewed after tag addition, MS and MS/MS were used for the mapping and sequencing of individual peptides, respectively. Peptide mixtures subjected to MS provided ion chromatograms (FIG. 3) that were used to compare the degree of ion overlap between the three examined molecules. Extensive ion overlap indicates that the peptides bound by these tailed and untailed B*1501 molecules were nearly identical.

Selected ions were then MS/MS sequenced in order to confirm that mapped ion overlaps indeed represented exact ligand matches through comparison of fragmentation patterns between the three molecules (Fig. 4). Approximately 60 peptides were chosen initially for MS/MS—ten from each fraction. Overall, fragmentation patterns were exact matches in a majority of the peptides examined (TABLE VI). Fragmentation patterns categorized as nonmatches resulted from a mixture of peptides present at the same mass to charge ratio, one or more of which was present in the tagged molecule and not apparent in the spectra of the same ion from B*1501. Of the sequence-level matches, ligands derived from HIS-tailed molecules more closely matched those derived from B*1501 than those eluted from FLAG-tailed

molecules. In total, 52/57 HIS peptides were exact matches, whereas 39/56 FLAG peptides were equivalent. Thus, the data indicates that the 6-HIS tag is less disruptive to endogenous peptide binding than is the FLAG-tag, although neither tag drastically altered the peptides bound by B*1501.

A handful of individual ligand sequences present in fractions of peptides eluted from all three molecules were determined by MS/MS. The two clearest sequences, AVRDISEASVF and QGLISRGYSY, demonstrate that tailed class I molecules indeed load endogenous peptide ligands. This supports the hypothesis that addition of a C-terminal tag does not abrogate the ability of the soluble HLA-B*1501 molecule to naturally bind endogenous peptides. Further, both peptide sequences closely matched those previously reported for B*1501 eluted peptides having a disparate N-terminus paired with a more conserved C-terminus consisting of either a phenylalanine or a tyrosine. Given the homologous Edman sequence, largely identical fragmentation patterns, and the peptide ligands shared between the three molecules, we conclude that addition of a C-terminal tag does not significantly alter the peptides bound by B*1501. Mapping and subtractively comparing eluted peptides is a direct means for identifying differences and similarities in the individual ligands bound by a class I HLA molecule. Indeed, subtractive comparisons demonstrate how overlapping ligands bind across closely related HLA-B15 subtypes as well as pointing out which ligands are unique to virus-infected cells. Direct comparative analyses of eluted peptide ligands is well suited for a number of purposes, not the least of which is viral and cancer CTL epitope discovery. Addition of a C-terminal epitope tag provides a feasible method for production and purification of class I molecules, and

therefore, peptide ligands in cell lines capable of sustaining viral infection or harboring neoplastic agents, as illustrated in FIG. 5. Direct peptide analysis from such lines should yield important information on host control of pathogenic elements as well as provide important building blocks for rational vaccine development.

The present invention further relates in particular to a novel method for detecting those peptide epitopes which distinguish the infected/tumor cell from the uninfected/non-tumor cell. The results obtained from the present inventive methodology cannot be predicted or ascertained indirectly; only with a direct epitope discovery method can the unique epitopes described herein be identified. Furthermore, only with this direct approach can it be ascertained that the source protein is degraded into potentially immunogenic peptide epitopes. Finally, this unique approach provides a glimpse of which proteins are uniquely up and down regulated in infected/tumor cells.

The utility of such HLA-presented peptide epitopes which mark the infected/tumor cell are three-fold. First, diagnostics designed to detect a disease state (i.e., infection or cancer) can use epitopes unique to infected/tumor cells to ascertain the presence/absence of a tumor/virus. Second, epitopes unique to infected/tumor cells represent vaccine candidates. Here, we describe epitopes which arise on the surface of cells infected with HIV. Such epitopes could not be predicted without natural virus infection and direct epitope discovery. The epitopes detected are derived from proteins unique to virus infected and tumor cells. These epitopes can be used for virus/tumor vaccine development and virus/tumor diagnostics. Third, the process indicates that particular proteins unique to virus infected cells are

found in compartments of the host cell they would otherwise not be found in. Thus, we identify uniquely upregulated or trafficked host proteins for drug targeting to kill infected cells.

The present invention describes, in particular, peptide epitopes unique to HIV infected cells. Peptide epitopes unique to the HLA molecules of HIV infected cells were identified by direct comparison to HLA peptide epitopes from uninfected cells.

As such, and only by example, the present method is shown to be capable of identifying: (1) HLA presented peptide epitopes, derived from intracellular host proteins, that are unique to infected cells but not found on uninfected cells, and (2) that the intracellular source-proteins of the peptides are uniquely expressed/processed in HIV infected cells such that peptide fragments of the proteins can be presented by HLA on infected cells but not on uninfected cells.

The method of the present invnetion also, therefore, describes the unique expression of proteins in infected cells or, alternatively, the unique trafficking and processing of normally expressed host proteins such that peptide fragments thereof are presented by HLA molecules on infected cells. These HLA presented peptide fragments of intracellular proteins represent powerful alternatives for diagnosing virus infected cells and for targeting infected cells for destruction (i.e., vaccine development).

A group of the host source-proteins for HLA presented peptide epitopes unique to HIV infected cells represent source-proteins that are uniquely expressed in cancerous cells. For example, through using the methodology of the present invention a peptide fragment of reticulocalbin is uniquely found

on HIV infected cells. A literature search indicates that the reticulocalbin gene is uniquely upregulated in cancer cells (breast cancer, liver cancer, colorectal cancer). Thus, the HLA presented peptide fragment of reticulocalbin which distinguishes HIV infected cells from uninfected cells can be inferred to also differentiate tumor cells from healthy non-tumor cells. Thus, HLA presented peptide fragments of host genes and gene products that distinguish the tumor cell and virus infected cell from healthy cells have been directly identified. The epitope discovery method of the present invention is also capable of identifiying host proteins that are uniquely expressed or uniquely processed on virus infected or tumor cells. HLA presented peptide fragments of such uniquely expressed or uniquely processed proteins can be used as vaccine epitopes and as diagnostic tools.

The methodology to target and detect virus infected cells may not be to target the virus-derived peptides. Rather, the methodology of the present invention indicates that the way to distinguish infected cells from healthy cells is through alterations in host encoded protein expression and processing. This is true for cancer as well as for virus infected cells. The methodology according to the present invention results in data which indicates without reservation that proteins/peptides distinguish virus/tumor cells from healthy cells.

Example of Comparative Ligand Mapping in Infected and Uninfected Cells Creation of Soluble Class I Construct

EBV-transformed cell lines expressing alleles of interest (particularly A*0201, B*0702, and Cw*0702) were grown and class I HLA typed through

the sequenced-based-typing methodology described in Turner et al. 1998, J. Immunol, 161 (3) 1406-13) and U.S. Patent No. 6,287,764 Hildebrand et al. both of which are expressly incorporated herein in their entirety by reference. Total RNA was 5pXI and 3pEI, producing a product lacking the cytoplasmic and transmembrane domains. Alternatively, a 3' primer encoding a hexa-histidine or FLAG epitope tag was placed on the C-terminus using the primers, 3pEIHIS or 3pEIFLAG (TABLE V). For the C-locus, a 5'primer was used encoding the Kozak consensus sequence. (Davis, et al. 1999. *J. Exp. Med.* 189: 1265-1274). Each construct was cut with the appropriate restriction endonculease (see TABLE V) and cloned into the mammalian expression vector pCDNA 3.1- (Invitrogen, Carlsbad, CA) encoding either a resistance gene for G418 sulfate or Zeocin (Invitrogen).

Transfection in Sup-T1 cells. Sup-T1 T cells were cultured in RPMI 1640 + 20% fetal calf serum at 37°C and 5% CO_2. Cells were split daily in order to maintain log-phase growth. Plasmid DNA was purified using either Qiagen Midi-prep kits (Qiagen, Santa Clarita) or Biorad Quantum Prep Midiprep Kit (Biorad, Hercules, Ca) according to the manufacturer's protocol and resuspended in sterile DNAse-free water. Cells were electroporated with 30 µgs of plasmid DNA at a voltage of 400 mV and a capacitance of 960 µF. Decay constants were monitored throughout electroporation and only transfections with decay times under 25 mS were carried through to selection. Selection was performed on day 4 post-transfection with .45 mg/mL Zeocin (Invitrogen) selective medium containing 30% fetal calf with the pH adjusted visually to just higher than neutral. Cells were resuspended in selective medium at 2 x 10^6 cells per ml, fed until they no longer turned the wells

yellow (using the pH indicated Phenol Red (Mediatech)), and allowed to sit until cells began to divide. After the appearance of active division, cells were slowly fed with selective medium until they reached medium (T-75) tissue culture flasks. Cells were then subcloned at limiting dilutions of .5, 1, and 1.5 cells per well in 96-well tissue culture plates. Cells were allowed to sit until well turned yellow, they were then gradually moved to 24 well plates and small (T-25) tissue culture flasks. Samples were taken for soluble class I ELISA, and the best producers of class I were frozen for later use at 5 x 10^6 cells/ml and stored at -135°C.

Soluble MHC class I ELISA. ELISAs were employed to test the concentration of the MHC class I/peptide complexes in cell culture supernatants. The monoclonal antibody W6/32 (ATCC, Manassas, Va) was used to coat 96-well Nunc Starwell Maxi-sorp plates (VWR, West Chester, Pa). One hundred µls of test sample containing class I was loaded into each well of the plate. Detection was with anti-βB2 microglobulin (light chain) antibody conjugated to horse-radish peroxidase followed by incubation with OPD (Sigma, St. Louis, MO). ELISA values were read by a SpectraMax 340 00A, Rom Version 2.04, February 1996, using the program Softmax Pro Version 2.2.1 from Molecular Devices. For determination of MHC class I complex in carboys prior to affinity purification (see below), each sample was tested in triplicate on at least 2 separate plates. Uninfected and infected harvest concentrations were read on the same plate and uninfected samples were brought to 1% Triton X 100 prior to loading on the ELISA plate. This was in an attempt to minimize variability in mass spectra generate due to large differences in the amount of peptide loaded onto affinity columns.

Full-length construct creation. Full-length constructs (in the pCDNA3.1-/G418 sulfate resistance vector) were created and transfected into the class I negative B-LCL 721.221 and T2. Both cell lines were cultured in RPMI-1640 + 10% fetal calf serum until growing at log phase. Cells were electroporated at 25 V and 960 µF capacitance. After 2 days, the cells were pelleted and resuspended in selective medium consisting of RPMI-1640 + 20% FCS + 1.5 mg/ml G418 sulfate (Mediatech, Herndon, Va). Cells were treated in the same manner as above (Sup-T1 transfection) after this point.

Cell pharm production. Eight liters of Sup-T1 soluble MHC class I transfectants cultured in roller bottles in RPMI-1640 + 15% FCS + 100 U penicillin/streptomycin were centrifuged for 10 min at 1100 X g. Supernatant was discarded and a total of 3 X 10^ 9 total cells were resuspended in 200 mls of conditioned medium. Infected cells were then added to a feed bottle and inoculated through the ECS feed pump of a Unisyn CP2500 cell pharm (Unisyn, Hopkington, MA) into 30 kD molecular-weight cut-off hollow-fiber bioreactors previously primed with RPMI-1640 containing 20% fetal calf serum. Cells were allowed to incubate overnight in the bioreactor at a temperature of 37°C and at a pH of 7.20 maintained automatically through CO_2 injection into the medium reservoir of the system. No new medium was introduced into the system during this time period and the ICS recirculation was maintained at a low value of 400 mls/minute. ECS feed was begun 12 hours post inoculation at a rate of 100 mls/day with 15% FCS supplemented RPMI-1640; ICS feed was likewise begun at a rate of 1 L/day. ECS recirculation was initiated at day 2 post-inoculation at a rate of 4 L/day. ECS and ICS samples were taken at 24-hour intervals and sHLA ELISAs (see

above) and glucose tests performed. ECS and ICS feed rates as well as ECS and ICS recirculation rates were adjusted based on increasing concentrations of sHLA in the harvest and decreasing levels of glucose in the ICS medium.

Virus production and infection HIV MN-1 production. HIV MN-1 cloned virus (Genbank Accession Number M17449) was thawed from frozen stock and used to infect 25 X 10^6 non-transfected Sup-T1 (Denny CT, et. al. 1986. Nature. 320:549.51, which is expressly incorporated herein in its entirety by reference) T cells using standard methods. Cells were cultured in RPMI-1640 +20% fetal bovine serum (MediaTech) for 5 days and observed for syncitia formation. Upon formation of syncitia, new cells were added in fresh RPMI-1640/20% FCS. Culture was continued for 5 more days when 100 mls of infected cells were removed. Supernatant was passed through a .45 um filter and cell-free virus was aliquotted and stored at -80°C. This process was continued until an appropriate amount of virus was harvested.

HIV-1 NL4-3 production. The infectious molecular clone pNL4-3 (Genbank Accession Number AF324493) was transformed into the *Esherichia coli* strain Top10F' (Invitrogen, Carlsbad, Ca). Plasmid DNA was midiprepped from transformed cells using either the Qiagen Midi Prep Kit (Qiagen, Santa Clarita, Ca) or the Biorad Quantum Prep Midiprep Kit (Biorad, Hercules, Ca) according to the manufacturer's instructions. Plasmid DNA was used to transfect 293T cells (GenHunter Corporation, Nashville, TN) using Roche's Fugene 6 reagent (Roche, Basel, Switzerland) following the manufacturer's protocol. Virus-containing supernatant was harvested at 24, 48, and 72 hours, clarified by centrifugation at 500 X g for 10 min, aliquotted, and stored at -80°C. Sup-T1 transfectants containing either soluble A*0201, B*0702, or

Cw*0702 were cocultured with virus resulting in high-titre virus. After 72 hours, infected cells were centrifuged at 1100 X g for 10 minutes. Supernatant containing cell-free virus was removed, passed through a .45 µm filter, aliquotted, and stored at −80°C. Virally-infected cells were resuspended in freeze medium (RPMI-1640 + 20% FCS + 10% DMSO) at approximately 6 X10^6 cells per ml and stored at −80°C.

Viral Titer Determination. One vial of frozen viral stock derived from either strain of HIV was thawed and used in a $TCID_{50}$ assay scored two ways: 1) wells containing at least 3 syncitia were considered positive or 2) wells containing over 50 ng/ml p24 antigen as determined by ELISA were considered positive. The $TCID_{50}$ was then calculated using the Spearman-Karber method (DAIDS Virology Manual for HIV Laboratories, Jan. 1997). The average of both scoring methods was used as the final titer of the virus. As a second means of viral titer monitoring, viral stock was used undiluted in a p24 ELISA (Beckman Coulter, Miami, Fl) in order to determine the ngs of p24 present in cell-free virus.

P24 ELISA. Determination of HIV p24 major core protein was determined by the commercially available Beckman Coulter p24 ELISA according to the manufacturer's instructions with the exceptions of the following modifications: samples were treated with 10% Triton-X 100 prior to removal from a BSL-3 facility, therefore the inactivation medium included in the kit was not used. Secondly, samples were serially diluted in water prior to use.

Hollow-fiber bioreactor culture of infected cells. All work including large-scale culture of HIV was performed in a Biosafety Level 3 Laboratory in accordance with guidelines set forth by the National Institutes of Health.

HIV MN-1 frozen viral stock aliquots were thawed and pooled to a 100 ml total volume, containing approximately 5.5 X 10^6 $TCID_{50}$'s. Eight liters of Sup-T1 soluble MHC class I transfectants cultured in roller bottles in RPMI-1640 + 15% FCS + 100 U penicillin/streptomycin were centrifuged for 10 min at 1100 X g. Supernatant was discarded and a total of 3 X 10^ 9 total cells were resuspended in 200 mls of conditioned medium. The 100 mls of cell-free HIV MN-1 was then added to the resuspended cells and incubated at 37°C in %5 CO_2 for 2 hours with gentle shaking every 20 minutes. Infected cells were then added to a feed bottle and inoculated through the ECS feed pump of a Unisyn CP2500 cell pharm (Unisyn, Hopkington, MA) into 30 kD molecular-weight cut-off hollow-fiber bioreactors previously primed with RPMI-1640 containing 20% fetal calf serum. Cells were allowed to incubate overnight in the bioreactor at a temperature of 37°C and at a pH of 7.20 maintained automatically through CO_2 injection into the medium reservoir of the system. No new medium was introduced into the system during this time period and the recirculation was maintained at a low value of 400 mls/minute. ECS feed was begun 12 hours post inoculation at a rate of 100 mls/day with 15% FCS supplemented RPMI-1640; ICS feed was likewise begun at a rate of 1 L/day. ECS and ICS samples were taken at 24-hour intervals, inactivated by addition of Triton-X 100 to 1%, and sHLA ELISAs, p24 ELISAs, and glucose tests performed as described above. ECS and ICS feed rates as well as ECS and ICS recirculation rates were adjusted based on increasing concentrations of sHLA in the harvest and decreasing levels of glucose in the ICS medium.

Soluble HLA purification. Soluble-HLA containing supernatant was removed in 1.9 L volumes from infected hollow-fiber bioreactors. Twenty-percent Triton-X 100 was sterilized and placed in 50 ml aliquots in 60 mls syringes; 2 syringes were injected into each 1.9 L harvest bottle as it was removed from the cell pharm, resulting in a final TX 100 percentage of 1%. Bottles were inverted gently several times to mix the TX 100 and stored at 4°C for a minimum of 1 week. After 1 week, harvest was centrifuged at 2000 X g for 10 minutes to remove cellular debris and pooled into 10 L carboys. An aliquot was then removed from the pooled, HIV-inactivated supernatant and used in a quantitative $TCID_{50}$ assay (as described above) and used to initiate a coculture with Sup-T1's. Only after demonstration of a completely negative coculture as well as $TCID_{50}$ were harvests removed from the BSL-3.

Class I/Peptide Production and Peptide Characterization Handling of MHC classI/peptide complexes from infected cells. Each 10L of cell pharm harvest was separated strictly on a temporal basis during the cell pharm run. (This was an attempt to assess any epitopic changes that might occur temporally during infection as opposed to those that might occur more globally.) Harvest was treated exactly as described above, except for the removal of a 2 ml aliquot for tests in both a $TCID_{50}$ assay and cell coculture assay to determine infectivity of the virus.

Affinity purification of infected and uninfected MHC class I complexes. Uninfected and infected harvest removed from CP2500 machines were treated in an identical manner post-removal from the cell pharm. Approximately 50 mgs total class I as measured by W6/32 ELISA (see above) were passed over a Pharmacia XK-50 (Amersham-Pharmacia Biotech, Piscataway, NJ) column

packed with 50 mls Sepharose Fast Flow 4B matrix (Amersham) coupled to W6/32 antibody. Bound class I complexes were washed first with 1 L 20 mM sodium phosphate wash buffer, followed by a wash with buffer containing the zwitterionic detergent Zwittergent 3-08 (Calbiochem, Merck KgaA, Darmstadt, Germany) at a concentration of 10 mM, plus NaCl at 50 mM, and 20 mM sodium phosphate. The zwittergent wash was monitored by UV absorption at a wavelength of 216nm for removal of Triton-X 100 hydrophobically bound to the peptide complexes. After 1 L of wash had passed over the column (more than a sufficient amount for the UV to return to baseline), zwittergent buffer was removed with 2 L of 20 mM sodium phosphate wash buffer. Peptides were eluted post wash with freshly made :2N acetic acid, pH 2.7.

Peptide isolation and separation. Post-elution, peptide-containing eluate fractions were brought up to 10% glacial acetic acid concentration through addition of 100% glacial acetic acid. Fractions were then pooled into a model 8050 stirred cell (Millipore, Bedford, MA) ultrafiltration device containing a 3 kD molecular-weight cutoff regenerated cellulose membrane (Millipore). The device was capped and tubing parafilmed to prevent leaks and placed in a 78°C water bath for 10 minutes. Post-removal, the peptide-containing elution buffer was allowed to cool to room temperature. The stirred cell was operated at a pressure of 55 psi under nitrogen flow. Peptides were collected in 50 ml conical centrifuge tubes (VWR, West Chester, Pa), flash frozen in super-cooled ethanol, and lyophilized to dryness. Peptides were resuspended either in 10% acetic acid or 10% acetonitrile. Peptides were purified through a first-round of HPLC on a Haisil C-18 column (Higgins Analytical, Moutain View, Ca), with an isocratic flow of 100% B (100% acetonitrile, .01% TFA)

for 40 minutes. Following elution, peptide-containing fractions were pooled, speed-vacuumed to dryness, and resuspended in 150 µls of 10% acetic acid. Two µgs of the base methyl violet were added to the peptide mixture in 10% acetic acid and this was loaded onto a Haisil C-18 column for fractionation. Peptides were fractionated by one of two methods, the latter resulting in increased peptide resolution. The first fractionation program was 2-10% B in 2 minutes, 10-60% B in 60 minutes, with 1 minute fraction collection. The second RP-HPLC gradient consisted of a 2-14% B in 2 minutes, 14-40% B in 60 minutes, 40-70% B in 20 minutes, with 1 minute fraction collection. Peptides eluting in a given fraction were monitored by UV absorbance at 216 nm. Separate but identical (down to the same buffer preparations) peptide purifications were done for each peptide-batch from uninfected and infected cells.

Mass-spectrometric mapping of fractionated peptides. Fractionated peptides were mapped by mass spectrometry to generate fraction-based ion maps. Fractions were speed-vacuumed to dryness and resuspended in 12 µls 50:50 methanol:water + .05% acetic acid. Two µls were removed and sprayed via nanoelectrospray (Protana, Odense, Denmark) into a Q-Star quadrupole mass spectrometer with a time-of-flight detector (Perseptive SCIEX, Foster City, Ca). Spectra were generated for masses in the range of 50-1200 amu using identical mass spectrometer settings for each fraction sprayed. Spectra were then base-line subtracted and analyzed using the programs BioMultiview version 1.5beta9 (Persceptive SCIEX) or BioAnalyst version 1.0 (Persceptive SCIEX). Spectra from the same fraction in uninfected/infected cells were manually aligned to the same mass range,

locked, and 15 amu increments visually assessed for the presence of differences in the ions represented by the spectra (for an example, see Hickman et al. 2000. *Human Immunology*. 61:1339-1346 which is expressly incorporated herein by reference). Ions were selected for MS/MS sequencing based on upregulations or downregulation of 1.5 fold over the same ion in the uninfected cells, or the presence or absence of the ion in infected cells. Ions were thus categorized into multiple categories prior to MS/MS sequencing.

Tandem mass-spectrometric analysis of selected peptides. Lists of ions masses corresponding to each of the following categories were generated: 1) upregulated in infected cells, 2) downregulated in infected cells, 3) present only in infected cells, 4) absent in infected cells, and 5) no change in infected cells. The last category was generally disregarded for MS/MS analysis and the first 4 categories were subjected to MS/MS sequencing on the Q-Star mass spectrometer. Peptide-containing fractions were sprayed into the mass spectrometer in 3 µl aliquots. All MS settings were kept constant except for the Q0 and Cad gas settings, which were varied to achieve the best fragmentation. Fragmentation patterns generated were interpreted manually and with the aid of BioMultiView version 1.5 beta 9. No sequencing algorithms were used for interpretation of data, however multiple web-based applications were employed to aid in peptide identification including: MASCOT (Perkins, DN et al. 1999. *Electrophoresis.* 20(18):3551-3567), Protein Prospector (Clauser K. R. et al. 1999. *Analytical Chemistry.* 71:2871), PeptideSearch (http://www.narrador.emblheidelberg.de/GroupPages/PageLink/Peptidesearchpage.html) and BLAST search

(http://www.ncbi.nlm.nih.gov/BLAST/).

Quality control of epitope changes. Multiple parameters were established before peptides identified in the above fashion were deemed "upregulated," "downregulated," etc. First, the peptide fractions before and after the fraction in which the peptide was identified were subjected to MS/MS at the same amu under the identical collision conditions employed in fragmentation of the peptide-of-interest and the spectra generated overlaid and compared. This was done to make sure that, in the unlikely event that the peptides had fractionated differently (even with methyl-violet base B standardization) there was not the presence of the peptide in an earlier or later fraction of the uninfected or infected peptides (and that the peptides had truly fractionated in an identical manner.) Secondly, the same amu that was used to identify the first peptide was then subjected to MS/MS in the alternate fraction (either infected or uninfected, whichever was opposite of the fraction in which the peptide was identified.) Spectra again were overlaid in order to prove conclusively that the fragmentation patterns did not match and thus the peptide was not present in the uninfected cells, or, in the case that the fragmentation patterns did match, that the peptides were upregulated in the infected cells. Finally, synthetic peptides were generated for each peptide identified. These peptides were resuspended in 10% acetic acid and RP-HPLC fractionated under the same conditions as employed for the original fractionation, ensuring that the peptide putatively identified had the same hydrophobicity as that of the ion MS/MS fragmented. This synthetic peptide was MS/MS fragmented under the same collision conditions as that of the ion, the spectra overlaid, and checked for an exact match with the original peptide

fragment.

Functional Analysis\Literature Searches. After identification of epitopes, literature searches were performed on source proteins to determine their function within the infected cell. Broad inferences can be made from the function of the protein. Source proteins were classified into groups according to functions inside the cell. Again, broad inferences can be made as to the groups of proteins that would be available for specific presentation solely on infected cells. Secondly, source proteins were scanned for other possible epitopes which may be bound by other MHC class I alleles. Peptide binding predictions (Parker, K. C., et. al. 1994. J. *Immunol.* 152:163) were employed to determine if other peptides presented from the source proteins were predicted to bind. Proteasomal prediction algorithms (A.K. Nussbaum, et. Al. 2001. *Immunogenetics* 53:87-94) were likewise employed to determine the likelihood of a peptide being created by the proteasome.

Sequence Identification. A discussion of the results seen with the application of this procedure is included using the peptide GPRTAALGLL as an example. Other examples and data obtained based on the methodology are listed in TABLE VII.

TABLE VII

ION	FRACTION	SEQUENCE	MW	OBS'D MW	SOURCE PROTEIN	START AA	ACCESSION #	CATEGORY	SEQ ID NO:
Peptides Identified on Infected cells that are not present on Uninfected Cells									
612.720	32INF	EQMFEDIISL	1223.582	1223.418	HIV MN-1; ENV	101		HIV-DERIVED	29
509.680	31INF	IPCLLISFL	1017.601	1017.334	CHOLINERGIC RECEPTOR, ALPHA-3 POLYPEPTIDE	250			30
469.180	31INF	STTAICATGL	936.466	936.360	UBIQUITIN-SPECIFIC PROTEASE	152	10720340		31
420.130	16INF	APAQNPEL	838.426	838.259	B-ASSOCIATED TRANSCRIPT PROTEIN 3 (BAT3)			MHC GENE PRODUCT	32
500.190	28INF	LVMAPRTVL	998.602	998.396	HLA-B HEAVY CHAIN LEADER SEQUENCE	2	4566550	MHC GENE PRODUCT	33
529.680	31INF	APF[INS]PADX	1057.388		UNKNOWN, CLOSE TO SEVERAL cDNA's			UNKNOWN	34
523.166	12INF	TPQSNRPVm	1044.500	1044.333	RNA POLYMERASE II POLYPEPTIDE A	527	4505939	RNA MACHINERY/BINDING PR	35
444.140	16INF	AARPATSTL	887.485	887.280	EUK. TRANSLATION INITIATION FACTOR 4	1073	Q04637	RNA MACHINERY/BINDING PR	36
470.650	16INF	MAMMAALMA	940.413	939.410	SPARC-LIKE PROTEIN	19	478522	TUMOR SUPPRESSOR GENE?	37
490.620	16INF	IATVDSYVI		979.240	TENASCIN-C (HEXABRACHION)	1823	13639246	TUMOR SUPPRESSOR GENE?	38
563.640	16INF	SPNQARAQAAL	1126.597	1126.364	POLYPYRIMIDINE TRACT-BINDING PROTEIN 1	141	131528	RNA MACHINERY/BINDING PR	39
	30INF	GPRTAALGLL	968.589	968.426	RETICULOCALBIN	4	4506457	TUMOR SUPPRESSOR GENE?	40
556.150	16INF	NPNQNKNVAL	1111.586	1111.300	ELAV (HuR)	188	4503551	RNA MACHINERY/BINDING PR	41
Peptides Identified on Uninfected cells that are not present on Infected cells									
	16UNINF	GSHSMRY			MHC CLASS I HEAVY CHAIN (could derive from multiple alleles, i.e., HLA-B*0702 or HLA-G, etc.)	variable	multiple	MHC Class I Product	42

The first step in identification of an epitope present only on uninfected cells is performing MS ion mapping. In this case, the reversed-phase HPLC fraction 30 obtained from HIV as disclosed hereinabove (which contains a fraction of the total class I peptides) was sprayed into the mass spectrometer and an ion spectrum created. FIG. 6 shows the sections of ion map in which an ion was first identified as upregulated. The ion at 484.74 can be seen to predominate in the upper map, which is the spectrum generated from peptides from the infected cells. One can also see that there are other peptides which differ in their intensities between the uninfected cells from one spectrum to another. After a peptide is initially identified, the area of the spectrum in which the peptide is found is zoomed in on in order to more fully see all the ions in the immediate area (FIG. 7). After zooming in on the area from 482-488 amu, the ion at 484.72 can be seen to only be present in the infected cells (which are seen in the spectrum on the top). A large difference such as this is not always seen, sometimes more minor differences are chosen for sequence determination. This ion, however, was considered an extremely good candidate for further analysis.

After identification of the ion, the next step in the process is to sequence the peptide by using tandem mass spectrometry. FIG. 8 shows the spectrum generated when the peptide is fragmented. These fragments are used to discern the amino acid sequence of the peptide. The sequence of this peptide was determined to be GPRTAALGLL. This peptide was isolated from infected HLA-B*0702 molecules. One early quality control step is examining the peptide's sequence to see if it fits the sequences that were previously shown to be presented by this molecule. B*0702 binds peptides

that have a G at their second position (P2) and an L as their C-terminal anchor. Based on this information, this sequence is likely to be a peptide presented by B*0702.

Descriptive characterization of peptide. Once the peptide sequence is obtained, information is gained on the source protein from which the peptide was derived in the cytosol of the infected cell. Initially, a BLAST search (available at the National Center for Biotechnology website) is done to provide protein information on the peptide. A BLAST search with the sequence GPRTAALGLL pulled up the protein reticulocalbin 2. After the source protein is known, information about the protein is ascertained first from the PubMed (again available at the National Center for Biotechnology website) and put into a format to which one can easily refer as seen in FIG. 9. All of the accession numbers for the protein, as well as the original description of the protein are included. This makes it easy to come back to the information for downstream use. Also, the protein sequence is copied, pasted, and saved as a text document for incorporation into later searches. The peptide is highlighted in the entire protein, giving some context as to where it is derived and how large the total protein is. This is the initial data gathering step post-sequence determination.

The next step in characterizing the ligand is doing literature searches on the source protein from which the peptide was derived. The protein is entered into the PubMed database and all entries with the word "reticulocalbin" are retrieved. FIG. 10 illustrates the listing that is done to summarize what has previously been described for this protein. It can be seen that for reticulocalbin, multiple articles have been published involving

this protein. The literature is summarized in a paragraph following the PubMed listings and put into the report. For reticulocalbin, some of the most interesting points are that it is an ER resident protein, which can lead to speculation on why it is presented on infected cells. Secondly, it has been previously found to be upregulated in several other types of cancers, such as breast and colorectal cancers. This again leads to speculation that this protein may be broadly applicable to treat more maladies than those caused by HIV. It is also determined whether or not this protein has been previously cited as interacting with/ or being interfered with by HIV. This was not seen for reticulocalbin and thus was not listed in the report, (although in some instances it is seen.) A broad understanding of the protein is gained through literature searches.

Predictive characterization of peptide. After the literature search, several secondary searches are performed. FIG. 11 illustrates the results of a peptide-binding algorithm performed using Parker's Prediction (which is described hereinabove). The entire source protein is used for input and the computer generates a list of peptides which are bound by the HLA allele chosen. In this case, B*0702 was chosen because that was the allele from which this peptide was derived. From the black arrow in the figure, it can be seen that the peptide sequenced by mass spectrometry is predicted to bind to HLA-B*0702 with a high affinity. Several other peptides are listed that are predicted to bind as well. FIG. 12 shows the same procedure being performed with the source peptide using another well-known search engine, SYPEITHI. (This engine can be found on the worldwide web using the URL: http://syfpeithi.bmi-heidelberg.com/Scripts/MHCServer.dll/EpPredict.htm.)

Again, the results from this search engine for B*0702 shows that this peptide is predicted to bind to HLA-B*0702 with a high affinity. Also, multiple other peptides are predicted to be derived from this source protein and bound. This prediction allows us to determine several things. First, we can tell if the peptide is predicted to be bound by previous algorithms. This allows us to know how well the programs work, and/or if other people could identify this peptide (if they had the source protein) from peptide binding algorithms. All of this information can be translated into increasing importance for the present inventive methodology not only for the peptide but also for the source protein itself.

After peptide-binding algorithms are performed, searches are done to determine whether the peptides would be created by the proteasome during normal processing of proteins into peptides. It should be strongly noted that multiple pathways for class I peptide loading are now being demonstrated and that the cleavage algorithms for human proteasomes are not well established by any means. While a positive result may indicate that the proteasome is largely responsible for cleavage, a negative result by no means indicates that the peptide is not presented in the class I molecule. FIG. 13 shows the results of the first proteasomal cleavage done for the source protein reticulocalbin using the cleavage predictor PaProC (available at URL http://www.paproc.de/). The epitope is outlined. By this prediction software, the peptide is not predicted to be cleaved by the normal proteasome. This may mean that in infected cells, alternative pathways of MHC class I presentation are being used, particularly in reference to the reticulocalbin peptide. This, in turn, may present novel methods for

therapeutics during viral infection. A second proteasomal cleavage search is also employed using the prediction software NetChop (available on the worldwide web) as seen in FIG. 14. By this prediction and other data from current literature in the field, the peptide would be created by the proteasome and cleaved to form the GPRTAALGLL identified.

A third round of analysis involves only the source protein. All other alleles are tested for peptide binding and lists of the highest binders generated. The proteasomal cleavage predictions are then referred to in order to elucidate how these peptides are generated. This information is useful for downstream testing of peptides and for determining whether or not this protein will be applicable for vaccine trials covering a broad range of HLA alleles. For reticulocalbin, multiple high-affinity peptides were demonstrated for differing HLA alleles (some examples of which are shown in FIG. 15) In this figure, several high affinity peptides deriving from reticulocalbin were identified for HLA-A*0201 and A*0101.

Quality control of sequence determination. There currently exists no direct means to score the quality of MS/MS sequence data. Once all descriptive and predictive steps are concluded, we return again to the original peptide sequence for quality control to ensure that the peptide is indeed what we have identified as the amino acid sequence and that the peptide is truly present only in infected cells. We employ these multiple steps so there is no doubt that the sequence is truly what we claim it to be before we move on to downstream applications involving the peptide.

Initially, we determine that the peptide is truly upregulated or present only in infected cells. For the reticulocalbin peptide, we determined that this

peptide was probably only present in infected cells. In order to make certain that the peptide was truly absent in the uninfected cells and that there was no chance that our RP-HPLC fractionation had differed (remembering that we use internal controls for our fractionation as well) we generated ion spectra using MS from the fractions before and after the one in which we identified the peptide. In the case of the reticulocalbin peptide, we identified the peptide in fraction 30, so we performed MS on fractions 29 and 31 (FIG. 16) In FIG. 16, it can be seen that there is no substantial peak at the m/z 484.72. This indicated that there was not differential fractionation and that the peptide truly was absent from uninfected cells. In the case that there was a peptide peak in one of the before or after fractions, we would then turn to MS/MS to determine whether this peak represented the ion we were characterizing or another ion with the same mass-to-charge ratio.

After determining that the peptide is not present in another fraction, MS/MS was preformed on the same m/z in the uninfected spectrum (in the same fraction) in order to conclusively prove that there is no peptide present with the same sequence in the uninfected cells. In FIG. 17 one can see that the fragmentation patterns produced under identical MS collision conditions are totally different. This illustrates the absence of the reticulocalbin peptide in the uninfected cells.

Finally, in order to conclusively prove that the peptide sequence is the same as that originally identified, we synthesize synthetic peptides consisting of the same amino acids as the peptide sequence identified from the MS/MS fragmentation pattern. For the reticulocalbin peptide (i.e. the ion in fraction 30 at 484.72) we synthesized the peptide "GPRTAALGLL." We

then took this peptide and did MS/MS on the peptide under identical conditions as previously used. FIG. 18 illustrates the spectrum generated from MS/MS of the endogenously loaded reticulocalbin peptide. Matching spectra, as seen here, are indicators that this peptide sequence is GPRTAALGLL as almost every amino acid combination will generate a completely different set of fragments, both in terms of production of fragments and in terms of intensity of those fragments present. FIG. 18 shows the MS/MS endogenous and synthetic "GPRTAALGLL" peptide under identical collision conditions. As can be seen, the MS/MS graphs are virtually identical.

In accordance with the present invention, one peptide ligand (i.e. "GPRTAALGLL") has been identified as being presented by the B*0702 class I MHC molecule in cells infected with the HIV MN-1 virus but not in uninfected cells. As one of ordinary skill in the art can appreciate the novelty and usefulness of the present methodology in directly identifying such peptide ligands and the importance such identification has for numerous therapeutic (vaccine development, drug targeting) and diagnostic tools. As such, numerous other peptide ligands have been uniquely identified in cells infected with HIV MN-1 (as opposed to uninfected cells_ and these results are summarized in TABLE VII. One of ordinary skill in the art given the present specification would be fully enabled to identify the "GPRTAALGLL" peptide ligand; as well as other uniquely presented peptide ligands found in cells infected with a microorganism of interest and/or tumorigenic cells.

As stated above, TABLE VII identifies the sequences of peptide ligands identified to date as being unique to HIV infected cells. Class I sHLA B*0702

was harvested for T cells infected and not infected with HIV. Peptide ligands were eluted from B*0702 and comparatively mapped on a mass spectrometer so that ions unique to infected cells were apparent. Ions unique to infected cells (and one ligand unique to uninfected cells) were subjected to mass spectrometric fragmentation for peptide sequencing. Column 1 indicates the ion selected for sequencing, column 2 is the HPLC fraction, column 3 is the peptide sequence, column 4 is the predicted molecular weight, column 5 is the molecular weight we found, column 6 is the source protein for the epitope sequenced, column 7 is where the epitope starts in the sequence of the source protein, column 8 is the accession number, and column 9 is a descriptor which briefly indicates what is known of that epitope and/or its source protein.

The methodology used herein is to use sHLA to determine what is unique to unhealthy cells as compared to healthy cells. Using sHLA to survey the contents of a cell provides a look at what is unique to unhealthy cells in terms of proteins that are processed into peptides. TABLE VII shows the utility of the method described herein for discovering epitopes and their source proteins which are unique to HIV infected cells. A detailed description of the peptide from Reticulocalbin is provided hereinabove. The other epitopes and corresponding source proteins described in TABLE VII were processed in the same manner as the reticulocalbin epitope and source protein were -- i.e. as described herein above. The data summarized in TABLE VII shows that the epitope discovery technique described herein is capable of identifying sHLA bound epitopes and their corresponding source proteins which are unique to infected/unhealthy cells.

Likewise, and as is shown in TABLE VII, peptide ligands presented in individual class I MHC molecules in an uninfected cell that are not presented by individual class I MHC molecules in an uninfected cell can also be identified. The peptide "GSHSMRY", for example, was identified by the method of the present invention as being an individual class I MHC molecule which is presented in an uninfected cell but not in an infected cell.

The utility of this data is at least threefold. First, the data indicates what comes out of the cell with HLA. Such data can be used to target CTL to unhealthy cells. Second, antibodies can be targeted to specifically recognize HLA molecules carrying the ligand described. Third, realization of the source protein can lead to therapies and diagnostics which target the source protein. Thus, an epitope unique to unhealthy cells also indicates that the source protein is unique in the unhealthy cell.

The methods of epitope discovery and comparative ligand mapping described herein are not limited to cells infected by a microorganism such as HIV. Unhealthy cells analyzed by the epitope discovery process described herein can arise from virus infection or also cancerous transformation. In addition, the status of an unhealthy cell can also be mimicked by transfecting a particular gene known to be expressed during viral infection or tumor formation. For example, particular genes of HIV can be expressed in a cell line as described (Achour, A., et al., AIDS Res Hum Retroviruses, 1994. 10(1): p. 19-25; and Chiba, M., et al., *CTL*. Arch Virol, 1999. 144(8): p. 1469-85, all of which are expressly incorporated herein by reference) and then the epitope discovery process performed to identify how the expression of the transferred gene modifies epitope presentation by sHLA. In a similar fashion, genes

known to be upregulated during cancer (Smith, E.S., et al., Nat Med, 2001. 7(8): p. 967-72, which is expressly incorporated herein by reference) can be transferred in cells with sHLA and epitope discovery then completed. Thus, epitope discovery with sHLA as described herein can be completed on cells infected with intact pathogens, cancerous cells or cell lines, or cells into which a particular cancer, viral, or bacterial gene has been transferred. In all these instances the sHLA described here will provide a means for detecting what changes in terms of epitope presentation and the source proteins for the epitopes.

Thus, in accordance with the present invention, there has been provided a methodology for epitope discovery and comparative ligand mapping which includes methodology for producing and manipulating Class I and Class II MHC molecules from gDNA that fully satisfies the objectives and advantages set forth herein above. Although the invention has been described in conjunction with the specific drawings, experimentation, results and language set forth herein above, it is evident that many alternatives, modifications, and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the invention.

All of the numerical and quantitative measurements set forth in this application (including in the examples and in the claims) are approximations.

The invention illustratively disclosed or claimed herein suitably may be practiced in the absence of any element which is not specifically disclosed or claimed herein. Thus, the invention may comprise, consist of, or consist essentially of the elements disclosed or claimed herein.

The following claims are entitled to the broadest possible scope consistent with this application. The claims shall not necessarily be limited to the preferred embodiments or to the embodiments shown in the examples.

What is claimed is:

1. A method of isolating peptide ligands for an individual class I molecule, comprising the steps of:

providing a cell line containing a construct that encodes an individual soluble class I molecule, the cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

culturing the cell line under conditions which allow for expression of the individual soluble class I molecules from the construct, such conditions also allowing for endogenous loading of a peptide ligand into the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having the endogenously loaded peptide ligands bound thereto; and

separating the peptide ligands from the individual soluble class I molecules.

2. The method of claim 1 wherein, in the step of providing a cell line containing a construct that encodes an individual soluble class I molecule, the construct further encodes a tag which is attached to the individual soluble class I molecule and aids in isolating the individual soluble class I molecule.

3. The method of claim 2, wherein the tag is selected from the group consisting of a HIS tail and a FLAG tail.

4. The method of claim 1 wherein, in the step of providing a cell line, the cell line is class I negative.

5. The method of claim 1 wherein, in the step of providing a cell line, the cell line expresses endogenous class I molecules.

6. The method of claim 1 wherein, in the step of providing a cell line, the cell line is infected with at least one of a microorganism, a gene from a microorganism, or a tumor gene.

7. The method of claim 6 wherein, the cell line is infected with HIV.

8. The method of claim 1 wherein, in the step of providing a cell line, the cell line is transformed such that the cell line is a tumor cell line.

9. The method of claim 1 wherein, in the step of providing a cell line containing a construct that encodes an individual soluble class I molecule, the cell line containing the construct that encodes the individual soluble class I molecule is produced by a method comprising the steps of:

obtaining genomic DNA or cDNA encoding at least one class I molecule;

identifying an allele encoding an individual class I molecule in the genomic DNA or cDNA;

PCR amplifying the allele encoding the individual class I molecule in a locus specific manner such that a PCR product produced therefrom encodes a truncated, soluble form of the individual class I molecule;

cloning the PCR product into an expression vector, thereby forming a construct that encodes the individual soluble class I molecule; and

transfecting the construct into a cell line.

10. The method of claim 9 wherein, in the step of providing a cell line containing a construct that encodes an individual soluble class I molecule, the construct further encodes a tag which is attached to the soluble class I molecule and aids in isolating the individual soluble class I molecule.

11. The method of claim 10, wherein the tag is selected from the group consisting of a HIS tail and a FLAG tail.

12. The method of claim 10, wherein the tag is encoded by a PCR primer utilized in the step of PCR amplifying the allele encoding the individual class I molecule.

13. The method of claim 10, wherein the tag is encoded by the expression vector into which the PCR product is cloned.

14. The method of claim 1, further comprising the step of identifying the peptide ligand.

15. The method of claim 14, further comprising the step of identifying a source protein from which the peptide ligand was obtained.

16. A peptide ligand for an individual class I molecule isolated by a method comprising the steps of:

providing a cell line containing a construct that encodes an individual soluble class I molecule, the cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

culturing the cell line under conditions which allow for expression of the individual soluble class I molecules from the construct, such conditions also allowing for endogenous loading of a peptide ligand into the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having the endogenously loaded peptide ligands bound thereto; and

separating the peptide ligands from the individual soluble class I molecules.

17. A peptide ligand for an individual class I molecule comprising SEQ ID NO:29.

18. The peptide ligand of claim 17, wherein the peptide ligand is obtained from HIV-1 MN, ENV.

19. A peptide ligand for an individual class I molecule comprising SEQ ID NO:30.

20. The peptide ligand of claim 19, wherein the peptide ligand is obtained from Cholinergic Receptor, Alpha-3 Polypeptide.

21. A peptide ligand for an individual class I molecule comprising SEQ ID NO:31.

22. The peptide ligand of claim 21, wherein the peptide ligand is obtained from Ubiquitin-Specific Protease.

23. A peptide ligand for an individual class I molecule comprising SEQ ID NO:32.

24. The peptide ligand of claim 23, wherein the peptide ligand is obtained from β-Associated Transcript Protein 3 (BAT3).

25. A peptide ligand for an individual class I molecule comprising SEQ ID NO:33.

26. The peptide ligand of claim 25, wherein the peptide ligand is obtained from HLA-B Heavy Chain Leader Sequence.

27. A peptide ligand for an individual class I molecule comprising SEQ ID NO:34.

28. A peptide ligand for an individual class I molecule comprising SEQ ID NO:35.

29. The peptide ligand of claim 28, wherein the peptide ligand is obtained from RNA Polymerase II Polypeptide A.

30. A peptide ligand for an individual class I molecule comprising SEQ ID NO:36.

31. The peptide ligand of claim 30, wherein the peptide ligand is obtained from EUK, Translation Initiation Factor 4.

32. A peptide ligand for an individual class I molecule comprising SEQ ID NO:37.

33. The peptide ligand of claim 32, wherein the peptide ligand is obtained from SPARC-1 Like Protein.

34. A peptide ligand for an individual class I molecule comprising SEQ ID NO:38.

35. The peptide ligand of claim 34, wherein the peptide ligand is obtained from Tenascin-C.

36. A peptide ligand for an individual class I molecule comprising SEQ ID NO:39.

37. The peptide ligand of claim 36, wherein the peptide ligand is obtained from Polypyrimidine Tract-Binding Protein 1.

38. A peptide ligand for an individual class I molecule comprising SEQ ID NO:40.

39. The peptide ligand of claim 38, wherein the peptide ligand is obtained from Reticulocalbin.

40. A peptide ligand for an individual class I molecule comprising SEQ ID NO:41.

41. The peptide ligand of claim 40, wherein the peptide ligand is obtained from ELAV (HuR).

42. A method for identifying at least one endogenously loaded peptide ligand that distinguishes an infected cell from an uninfected cell, comprising the steps of:

providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the uninfected cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

infecting a portion of the uninfected cell line with at least one of a microorganism, a gene from a microorganism or a tumor gene, thereby providing an infected cell line;

culturing the uninfected cell line and the infected cell line under conditions which allow for expression of individual soluble class I molecules from the construct, such conditions also allowing for endogenous loading of a peptide ligand in the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having the endogenously loaded peptide ligands bound thereto from the uninfected cell line and the infected cell line;

separating the endogenously loaded peptide ligands from the individual soluble class I molecules from the uninfected cell line and separating the endogenously loaded peptide ligands from the individual soluble class I molecules from the infected cell line;

isolating the endogenously loaded peptide ligands from the uninfected cell line and the endogenously loaded peptide ligands from the infected cell line;

comparing the endogenously loaded peptide ligands isolated from the infected cell line to the endogenously loaded peptide ligands isolated from the uninfected cell line; and

identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the infected cell line that is not presented by the individual soluble class I molecule on the uninfected cell line.

43. The method of claim 42 wherein, in the step of providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the construct further encodes a tag which is attached to the individual soluble class I molecule and aids in isolating the individual soluble class I molecule.

44. The method of claim 42 wherein, the uninfected cell line is class I negative.

45. The method of claim 42 wherein, the uninfected cell line expresses endogenous class I molecules.

46. The method of claim 42 wherein, in the step of providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the uninfected cell line containing the construct that encodes the individual soluble class I molecule is produced by a method comprising the steps of:

obtaining genomic DNA or cDNA encoding at least one class I molecule;

identifying an allele encoding an individual class I molecule in the genomic DNA or cDNA;

PCR amplifying the allele encoding the individual class I molecule in a locus specific manner such that a PCR product produced therefrom encodes a truncated, soluble form of the individual class I molecule;

cloning the PCR product into an expression vector, thereby forming a construct that encodes the individual soluble class I molecule; and

transfecting the construct into an uninfected cell line.

47. The method of claim 46 wherein, in the step of providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the construct further encodes a tag which is attached to the soluble class I molecule and aids in isolating the individual soluble class I molecule.

48. The method of claim 47, wherein the tag is selected from the group consisting of a HIS tail and a FLAG tail.

49. The method of claim 47, wherein the tag is encoded by a PCR primer utilized in the step of PCR amplifying an allele encoding the individual class I molecule.

50. The method of claim 47, wherein the tag is encoded by the expression vector into which the PCR product is cloned.

51. The method of claim 42, further comprising the step of identifying a source protein from which the at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the infected cell line and not presented by the individual soluble class I molecule on the uninfected cell line is obtained.

52. The method of claim 42 wherein, in the step of identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the infected cell line but not on the uninfected cell line, the at least one endogenously loaded peptide ligand is obtained from a protein encoded by at least one of the microorganism, the gene from a microorganism or the tumor gene with which the cell line was infected to form the infected cell line.

53. The method of claim 42 wherein, in the step of identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the infected cell line but not on the uninfected cell line, the at least one endogenously loaded peptide ligand is obtained from a protein encoded by the uninfected cell line.

54. The method of claim 53, wherein the protein encoded by the uninfected cell line from which the at least one endogenously loaded peptide ligand is obtained has increased expression in a tumor cell line.

55. The method of claim 42 wherein, in the step of infecting a portion of the uninfected cell line, the portion of the uninfected cell line is infected with HIV.

56. A method for identifying at least one endogenously loaded peptide ligand that distinguishes an infected cell from an uninfected cell, comprising the steps of:

providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the uninfected cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

infecting a portion of the uninfected cell line with at least one of a microorganism, a gene from a microorganism or a tumor gene, thereby providing an infected cell line;

culturing the uninfected cell line and the infected cell line under conditions which allow for expression of the individual soluble class I molecules from the construct, such conditions also allowing for endogenous loading of a peptide ligand into the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having the endogenously loaded peptide ligands bound thereto from the uninfected cell line and the infected cell line;

separating the endogenously loaded peptide ligands from the individual soluble class I molecules from the uninfected cell line and separating the endogenously loaded peptide ligands from the individual soluble class I molecules from the infected cell line;

isolating the endogenously loaded peptide ligands from the uninfected cell line and the endogenously loaded peptide ligands from the infected cell line;

comparing the endogenously loaded peptide ligands isolated from the uninfected cell line to the endogenously loaded peptide ligands isolated from the infected cell line; and

identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the uninfected cell line that is not presented by the individual soluble class I molecule on the infected cell line.

57. The method of claim 56 wherein, in the step of providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the construct further encodes a tag which is attached to the individual soluble class I molecule and aids in isolating the individual soluble class I molecule.

58. The method of claim 56, wherein the uninfected cell line is class I negative.

59. The method of claim 56, wherein the uninfected cell line expresses endogenous class I molecules.

60. The method of claim 56 wherein, in the step of providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the uninfected cell line containing the construct that encodes the individual soluble class I molecule is produced by a method comprising the steps of:

obtaining genomic DNA or cDNA encoding at least one class I molecule;

identifying an allele encoding an individual class I molecule in the genomic DNA or cDNA:

PCR amplifying the allele encoding the individual class I molecule in a locus specific manner such that a PCR product produced therefrom encodes a truncated, soluble form of the individual class I molecule;

cloning the PCR product into an expression vector, thereby forming a construct that encodes the individual soluble class I molecule; and

transfecting the construct into an uninfected cell line.

61. The method of claim 60 wherein, in the step of providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the construct further encodes a tag which is attached to the individual soluble class I molecule and aids in isolating the individual soluble class I molecule.

62. The method of claim 61, wherein the tag is selected from the group consisting of a HIS tail and a FLAG tail.

63. The method of claim 61, wherein the tag is encoded by a PCR primer utilized in the step of PCR amplifying the allele encoding the individual class I molecule.

64. The method of claim 62 wherein the tag is encoded by the expression vector into which the PCR product is cloned.

65. The method of claim 56, further comprising the step of identifying a source protein from which the at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the uninfected

cell line and not presented by the individual soluble class I molecule on the infected cell line is obtained.

66. The method of claim 56 wherein, in the step of infecting a portion of the uninfected cell line, the portion of the uninfected cell line is infected with HIV.

67. An endogenously loaded peptide ligand presented by an individual class I molecule on an infected cell but not on an uninfected cell.

68. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:29.

69. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:30.

70. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:31.

71. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:32.

72. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:33.

73. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:34.

74. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:35.

75. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:36.

76. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:37.

77. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:38.

78. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:39.

79. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:40.

80. The endogenously loaded peptide ligand of claim 67, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:41.

81. A self protein capable of being naturally processed into at least one peptide fragment, wherein the at least one peptide fragment is endogenously loaded in and presented by a class I molecule on an infected cell.

82. The self protein of claim 81 wherein the at least one peptide fragment is not endogenously loaded in and presented by a class I molecule on an uninfected cell.

83. A peptide ligand presented by an individual class I molecule on an infected cell but not on an uninfected cell, the peptide ligand identified by a method comprising the steps of:

providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

infecting a portion of the uninfected cell line with at least one of a microorganism, a gene from a microorganism or a tumor gene, thereby providing an infected cell line;

culturing the uninfected cell line and the infected cell line under conditions which allow for expression of the individual soluble class I molecules from the construct, such conditions also allowing for endogenous loading of a peptide ligand in the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having the endogenously loaded peptide ligands bound thereto from the uninfected cell line and the infected cell line;

separating the endogenously loaded peptide ligands from the individual soluble class I molecules from the uninfected cell and the endogenously loaded peptide ligands from the individual soluble class I molecules from the infected cell;

isolating the endogenously loaded peptide ligands from the uninfected cell line and the endogenously loaded peptide ligands from the infected cell line;

comparing the endogenously loaded peptide ligands isolated from the infected cell line to the endogenously loaded peptide ligands isolated from the uninfected cell line; and

identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the infected cell line that is not presented by the individual soluble class I molecule on the uninfected cell line.

84. The peptide ligand of claim 83 wherein, in the step of providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the uninfected cell line containing the construct that encodes the individual soluble class I molecule is produced by a method comprising the steps of:

obtaining genomic DNA or cDNA encoding at least one class I molecule;

identifying an allele encoding an individual class I molecule in the genomic DNA or cDNA;

PCR amplifying the allele encoding the individual class I molecule in a locus specific manner such that a PCR product produced therefrom encodes a truncated, soluble form of the individual class I molecule;

cloning the PCR product into an expression vector, thereby forming a construct that encodes the individual soluble class I molecule; and

transfecting the construct into an uninfected cell line.

85. The peptide ligand of claim 84, wherein the construct further encodes a tag which is attached to the individual soluble class I molecule and aids in isolating the individual soluble class I molecule.

86. The peptide ligand of claim 85, wherein the tag is selected from the group consisting of a HIS tail and a FLAG tail.

87. The peptide ligand of claim 85, wherein the tag is encoded by a PCR primer utilized in the step of PCR amplifying the allele encoding the individual class I molecule.

88. The peptide ligand of claim 85, wherein the tag is encoded by the expression vector into which the PCR product is cloned.

89. The peptide ligand of claim 83, wherein the at least one endogenously loaded peptide ligand is obtained from a protein encoded by at least one of the microorganism, the gene from the microorganism or the tumor gene with which the portion of the uninfected cell line is infected to form the infected cell line.

90. The peptide ligand of claim 83, wherein the at least one endogenously loaded peptide ligand is obtained from a protein encoded by the uninfected cell line.

91. The peptide ligand of claim 83, wherein the portion of the uninfected cell line is infected with HIV.

92. A source protein from which the peptide ligand of claim 83 is obtained.

93. A peptide ligand endogenously loaded in an individual class I molecule and presented by the individual class I molecule on an uninfected cell.

94. The peptide ligand of claim 93 wherein the peptide ligand is not endogenously loaded in the individual class I molecule and presented by the individual class I molecule on an infected cell.

95. A self protein capable of being processed into at least one peptide fragment, wherein the at least one peptide fragment is endogenously loaded in an individual class I molecule and presented by the individual class I molecule on an uninfected cell.

96. The self protein of claim 95 wherein the at least one peptide fragment is not endogenously loaded in the individual class I molecule and presented by the individual class I molecule on an infected cell.

97. A peptide ligand endogenously loaded in an individual class I molecule and presented by the individual class I molecule on an uninfected cell but not on an infected cell, the peptide ligand identified by a method comprising the steps of:

providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the uninfected cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

infecting a portion of the uninfected cell line with at least one of a microorganism, a gene from a microorganism or a tumor gene, thereby providing an infected cell line;

culturing the uninfected cell line and the infected cell line under conditions which allow for expression of the individual soluble class I molecules, such conditions also allowing for endogenous loading of a peptide ligand in the antigen binding groove of each

individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having the endogenously loaded peptide ligands bound thereto from the uninfected cell line and the infected cell line;

separating the endogenously loaded peptide ligands from the individual soluble class I molecules from the uninfected cell line and the infected cell line;

isolating the endogenously loaded peptide ligands from the uninfected cell line and the endogenously loaded peptide ligands from the infected cell line;

comparing the endogenously loaded peptide ligands isolated from the uninfected cell line to the endogenously loaded peptide ligands isolated from the infected cell line; and

identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the uninfected cell line that is not presented by the individual soluble class I molecule on the infected cell line.

98. The peptide ligand of claim 97 wherein the construct further encodes a tag which is attached to the individual soluble class I molecule and aids in isolating the individual soluble class I molecule.

99. The peptide ligand of claim 97 wherein the uninfected cell line containing the construct that encodes the individual soluble class I molecule is produced by a method comprising the steps of:

obtaining genomic DNA or cDNA encoding at least one class I molecule;

identifying an allele encoding an individual class I molecule in the genomic DNA or cDNA;

PCR amplifying the allele encoding the individual class I molecule in a locus specific manner such that a PCR product produced therefrom encodes a truncated, soluble form of the individual class I molecule;

cloning the PCR product into an expression vector, thereby forming a construct that encodes an individual soluble class I molecule; and

transfecting the construct into an uninfected cell line.

100. The peptide ligand of claim 99, wherein the construct further encodes a tag which is attached to the individual soluble class I molecule and aids in isolating the individual soluble class I molecule.

101. The peptide ligand of claim 99, wherein the tag is selected from the group consisting of a HIS tail and a FLAG tail.

102. The peptide ligand of claim 99, wherein the tag is encoded by a PCR primer utilized in the step of PCR amplifying an allele encoding the individual class I molecule.

103. The method of claim 99, wherein the tag is encoded by the expression vector into which the PCR product is cloned.

104. A source protein from which the peptide ligand of claim 97 is obtained.

105. A method for identifying a self protein that is processed into at least one peptide fragment, wherein the at least one peptide fragment is

endogenously loaded in an individual class I molecule and presented by the individual class I molecule on an infected cell but not on an uninfected cell, the method comprising the steps of:

providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the uninfected cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

infecting a portion of the uninfected cell line with at least one of a microorganism, a gene from a microorganism or a tumor gene, thereby providing an infected cell line;

culturing the uninfected cell line and the infected cell line under conditions which allow for expression of the individual soluble class I molecules, such conditions also allowing for endogenous loading of a peptide ligand in the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having endogenously loaded peptide ligands bound thereto from the uninfected cell line and the infected cell line;

separating the endogenously loaded peptide ligands from the individual soluble class I molecules from the uninfected cell line and the infected cell line;

isolating the endogenously loaded peptide ligands from the uninfected cell line and the endogenously loaded peptide ligands from the infected cell line;

comparing the endogenously loaded peptide ligands isolated from the infected cell line to the endogenously loaded peptide ligands isolated from the uninfected cell line:

identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the infected cell line that is not presented by the individual soluble class I molecule on the uninfected cell line;

determining the source protein from which the at least one endogenously loaded peptide ligand is obtained; and

identifying the source protein as a self protein if the source protein is not encoded by the microorganism, gene from a microorganism or tumor gene with which the infected cell line is infected but is encoded by the uninfected cell line.

106. A method for identifying a self protein that is processed into at least one peptide fragment, wherein the at least one peptide fragment is endogenously loaded in an individual class I molecule presented by the individual class I molecule on an uninfected cell but not on an infected cell, the method comprising the steps of:

providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the uninfected cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

infecting a portion of the uninfected cell line with at least one of a microorganism, a gene from a microorganism or a tumor gene, thereby providing an infected cell line;

culturing the uninfected cell line and the infected cell line under conditions which allow for expression of the individual soluble class I molecules, such conditions also allowing for endogenous loading of a peptide ligand in the antigen binding groove of each

individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having endogenously loaded peptide ligands bound thereto from the uninfected cell line and from the infected cell line;

separating the endogenously loaded peptide ligands from the soluble class I molecules from the uninfected cell line and from the infected cell line;

isolating the endogenously loaded peptide ligands from the uninfected cell line and the endogenously loaded peptide ligands from the infected cell line;

comparing the endogenously loaded peptide ligands isolated from the uninfected cell line to the endogenously loaded peptide ligands isolated from the infected cell line;

identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the uninfected cell line that is not presented by the individual soluble class I molecule on the infected cell line; and

determining the source protein from which the at least one endogenously loaded peptide ligand is obtained.

107. A kit for identifying endogenously loaded peptide ligands for an individual class I molecule, comprising:

a cell line containing a construct that encodes an individual soluble class I molecule, the cell line capable of naturally processing proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules such that when the cell line is cultured under conditions which allow for expression of the

individual soluble class I molecule, a peptide ligand is endogenously loaded into the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell.

108. The kit of claim 107 wherein the cell line is produced by a method comprising the steps of:

obtaining genomic DNA or cDNA encoding at least one class I molecule;

identifying an allele encoding an individual class I molecule in the genomic DNA or cDNA;

PCR amplifying the allele encoding the individual class I molecule in a locus specific manner such that a PCR product produced therefrom encodes a truncated, soluble form of the individual class I molecule;

cloning the PCR product into an expression vector, thereby forming a construct that encodes an individual soluble class I molecule; and

transfecting the construct into an uninfected cell line.

109. A method for detecting a disease state, comprising the step of:

providing means for detecting an endogenously loaded peptide ligand in an individual class I molecule, wherein the endogenously loaded peptide ligand is presented by an individual class I molecule on an infected cell but not on an uninfected cell.

110. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:29.

111. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:30.

112. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:31.

113. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:32.

114. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:33.

115. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:34.

116. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:35.

117. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:36.

118. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:37.

119. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:38.

120. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:39.

121. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:40.

122. The method of claim 109, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:41.

123. The method of claim 109, wherein the disease state is HIV infection.

124. A method for detecting a disease state, comprising the step of:
providing means for detecting an endogenously loaded peptide ligand in an individual class I molecule, wherein the endogenously loaded peptide ligand is presented by the individual class I molecule on an uninfected cell but not on an infected cell.

125. A kit for detecting a disease state, comprising:
means for detecting an endogenously loaded peptide ligand in an individual class I molecule, wherein the endogenously loaded peptide ligand is presented by the individual class I molecule on an infected cell but not on an uninfected cell.

126. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:29.

127. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:30.

128. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:31.

129. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:32.

130. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:33.

131. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:34.

132. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:35.

133. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:36.

135. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:37.

136. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:38.

137. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:39.

138. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:40.

139. The method of claim 125, wherein the endogenously loaded peptide ligand comprises SEQ ID NO:41.

140. A kit for detecting a disease state, comprising:

means for detecting a peptide ligand endogenously loaded in an individual class I molecule, wherein the peptide ligand is presented by the individual class I molecule on an uninfected cell but not on an infected cell.

141. A method of identifying a motif for endogenously loaded peptide ligands presented by an individual class I molecule, comprising the steps of:

providing a cell line containing a construct that encodes an individual soluble class I molecule, the cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

culturing the cell line under conditions which allow for expression of the individual soluble class I molecule from the construct, such conditions also allowing for endogenous loading of the peptide ligand into the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having endogenously loaded peptide ligands bound thereto;

separating the endogenously loaded peptide ligands from the individual soluble class I molecules, thereby forming a pool of endogenously loaded peptide ligands; and

sequencing the pool of endogenously loaded peptide ligands and deriving a motif for the endogenously loaded peptide ligands based on dominant occurrences of particular amino acids at specific positions of the endogenously loaded peptide ligands.

142. A method of targeting a compound to an infected cell, comprising the steps of:

providing an uninfected cell line containing a construct that encodes an individual soluble class I molecule, the cell line being able to naturally process proteins into peptide ligands capable of being loaded into antigen binding grooves of class I molecules;

infecting a portion of such cell line with at least one of a microorganism, a gene from a microorganism or a tumor gene, thereby providing an infected cell line;

culturing the uninfected cell line and the infected cell line under conditions which allow for expression of the individual soluble class I molecule, such conditions also allowing for endogenous loading of a peptide ligand in the antigen binding groove of each individual soluble class I molecule prior to secretion of the individual soluble class I molecules from the cell;

isolating the secreted individual soluble class I molecules having endogenously loaded peptide ligands bound thereto from the uninfected cell line and the infected cell line;

separating the endogenously loaded peptide ligands from the individual soluble class I molecules from the uninfected cell line and the infected cell line;

isolating the endogenously loaded peptide ligands from the uninfected cell line and the endogenously loaded peptide ligands from the infected cell line;

comparing the endogenously loaded peptide ligands isolated from the infected cell line to the endogenously loaded peptide ligands isolated from the uninfected cell line; and

identifying at least one endogenously loaded peptide ligand presented by the individual soluble class I molecule on the infected cell line that is not presented by the individual soluble class I molecule on the uninfected cell line; and

targeting a compound to a cell having the at least one endogenously loaded peptide ligand presented by the individual class I molecule on a surface of the cell.

143. The method of claim 142 wherein, in the step of targeting a compound to a cell, the compound is a drug.

144. The method of claim 142 wherein, in the step of targeting a compound to a cell, the compound is an antibody.

145. The method of claim 142 wherein, in the step of targeting a compound to a cell, the compound specifically recognizes a complex formed of the at least one endogenously loaded peptide ligand and the individual class I molecule.

146. A method of isolating ligands, comprising the steps of:
optionally providing a cell line;
optionally culturing the cell line;
isolating molecules having ligands bound thereto; and
optionally separating the ligands from the molecules.

147. A ligand isolated by a method comprising the steps of:
optionally providing a cell line;
optionally culturing the cell line;
isolating molecules having ligands bound thereto; and
optionally separating the ligands from the molecules.

148. A method for identifying at least one ligand, comprising the steps of:
optionally providing a cell line;
optionally infecting a portion of the cell line;
optionally culturing the cell line;
optionally isolating molecules having ligands bound thereto;
optionally separating the ligands from the molecules;
optionally isolating the ligands; and
identifying at least one ligand.

149. A method for identifying at least one ligand, comprising the steps of:
optionally providing an uninfected cell line;

optionally infecting a portion of the uninfected cell line;

optionally culturing the uninfected cell line and the infected cell line;

optionally isolating molecules having ligands bound thereto from the uninfected cell line and the infected cell line;

optionally separating the ligands from the molecules from the uninfected cell line and separating the ligands from the molecules from the infected cell line;

optionally isolating the ligands from the uninfected cell line and the ligands from the infected cell line;

optionally comparing the ligands isolated from the uninfected cell line to the ligands isolated from the infected cell line; and

identifying at least one ligand presented by the molecule on the uninfected cell line that is not presented by the molecule on the infected cell line.

150. A ligand presented by an individual class I molecule on an infected cell but optionally not on an uninfected cell.

151. A protein capable of being processed into at least one peptide fragment, wherein the at least one peptide fragment is loaded in and/or presented by a class I molecule.

152. A ligand identified by a method comprising the steps of:

optionally infecting a portion of the cell line;

optionally culturing the cell line;

optionally isolating molecules having ligands bound thereto;

optionally separating the ligands from the molecules;

optionally isolating the ligands; and

identifying at least one ligand.

153. A ligand loaded in a molecule and presented by the molecule on a cell.

154. A protein capable of being processed into at least one peptide fragment, wherein the at least one peptide fragment is loaded in an individual class I molecule and is optionally presented by the individual class I molecule on an uninfected cell.

155. A ligand identified by a method comprising the steps of:

optionally providing an uninfected cell line;

optionally infecting a portion of the uninfected cell line;

optionally culturing the uninfected cell line and the infected cell line;

optionally isolating molecules having ligands bound thereto from the uninfected cell line and the infected cell line;

optionally separating the ligands from the molecules from the uninfected cell line and the infected cell line;

optionally isolating the ligands from the uninfected cell line and the

optionally comparing the ligands isolated from the uninfected cell line to the ligands isolated from the infected cell line; and

identifying at least one peptide ligand presented by the molecule on the uninfected cell line that is not presented by the molecule on the infected cell line.

156. A method for identifying a self protein, the method comprising the steps of:

optionally providing a cell line;

optionally infecting a portion of the cell line;

optionally culturing the cell line;

optionally isolating molecules having ligands bound thereto;

optionally separating the ligands from the molecules;

optionally isolating the ligands;

optionally identifying at least one ligand;

determining a source protein from which the at least one ligand is obtained; and

optionally identifying the source protein as the self protein.

157. A method for identifying a self protein, the method comprising the steps of:

optionally providing an uninfected cell line;

optionally infecting a portion of the uninfected cell line;

optionally culturing the uninfected cell line and the infected cell line;

optionally isolating molecules having ligands bound thereto from the uninfected cell line and from the infected cell line;

optionally separating the ligands from the molecules from the uninfected cell line and from the infected cell line;

optionally isolating the ligands from the uninfected cell line and the ligands from the infected cell line;

optionally comparing the ligands isolated from the uninfected cell line to the ligands isolated from the infected cell line;

optionally identifying at least one ligand presented by the molecule on the uninfected cell line that is not presented by the molecule on the infected cell line; and

determining a source protein from which the at least one ligand is obtained.

158. A kit for identifying ligands, comprising:

a cell line containing a construct that encodes a class I molecule.

159. A method for detecting a disease state, comprising the step of:

providing means for detecting a ligand in a molecule.

160. A method for detecting a disease state, comprising the step of:

providing means for detecting a ligand in a molecule, wherein the ligand optionally is presented by the molecule on an uninfected cell but not

161. A kit for detecting a disease state, comprising:

means for detecting a ligand in a molecule.

162. A kit for detecting a disease state, comprising:

means for detecting a ligand, wherein the ligand optionally is presented by an individual class I molecule on an uninfected cell but not on an infected cell.

163. A method of identifying a motif for ligands, comprising the steps of:

optionally providing a cell line;

optionally culturing the cell line;

optionally isolating molecules having ligands bound thereto;

optionally separating the ligands from the molecules, thereby forming a pool of ligands; and

sequencing the pool of ligands and deriving a motif for the ligands.

164. A method of targeting a compound to a cell, comprising the steps of:

optionally providing a cell line;

optionally infecting a portion of such cell line;

optionally culturing the cell line;

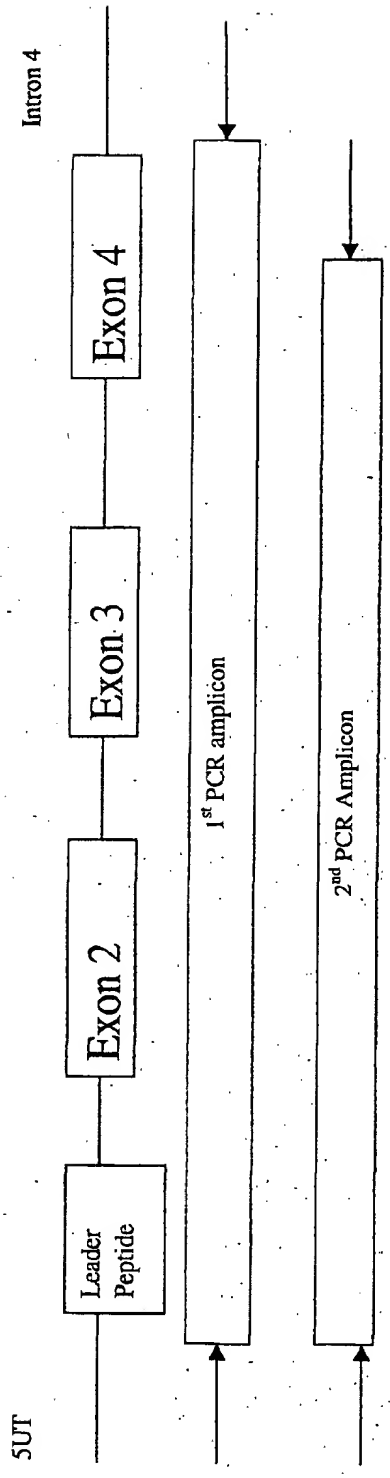
optionally isolating molecules having endogenously loaded peptide ligands bound thereto;

optionally separating the ligands from the molecules;

optionally identifying at least one ligand; and

targeting a compound to a cell having the at least one ligand.

Fig 1.


Intron 4
Exon 4
Exon 3
Exon 2
Leader Peptide
5UT
1st PCR amplicon
2nd PCR Amplicon

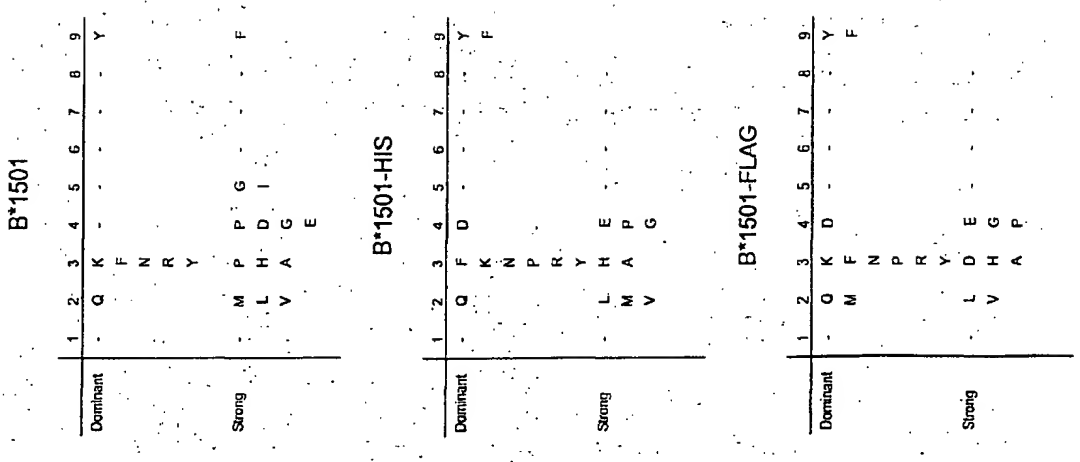

B*1501
1 2 3 4 5 6 7 8 9
Dominant - Q K - - - - - Y
F
N
R
Y
Strong - M P P G - - - F
L H D I
V A G
E
B*1501-HIS
1 2 3 4 5 6 7 8 9
Dominant - Q F D - - - - Y
F
K
N
P
R
Y
Strong - L H E - - -
M A P
V G
B*1501-FLAG
1 2 3 4 5 6 7 8 9
Dominant - Q K D - - - - Y
M F F
N
P
R
Y
Strong - L D E - - -
V H G
A P

FIGURE 2

FIGURE 3

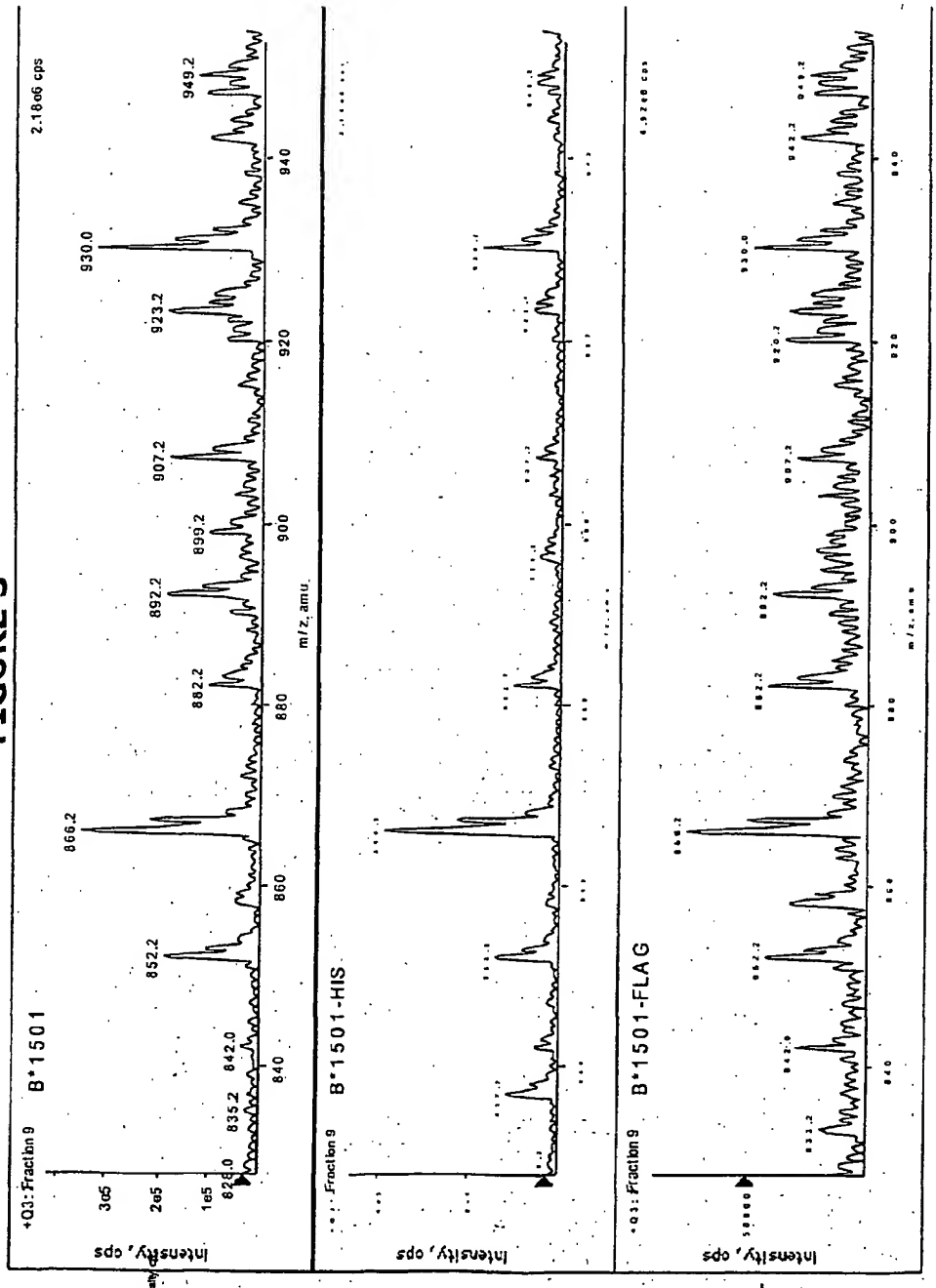

B*1501
B*1501-HIS
B*1501-FLAG
Intensity, cps
m/z, amu
2.18e6 cps
866.2
852.2
842.0
835.2
882.2
892.2
899.2
907.2
923.2
930.0
949.2

FIGURE 4

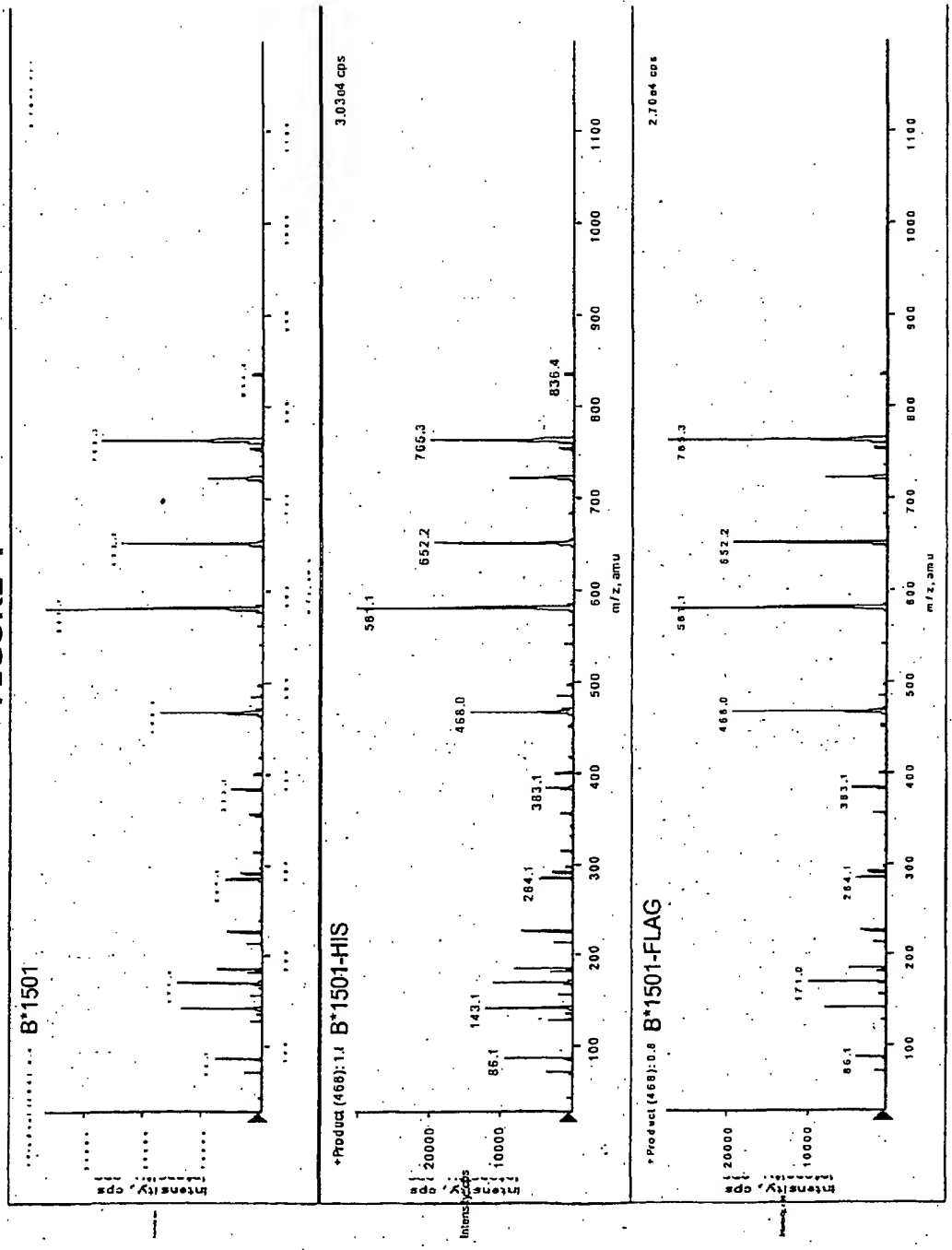
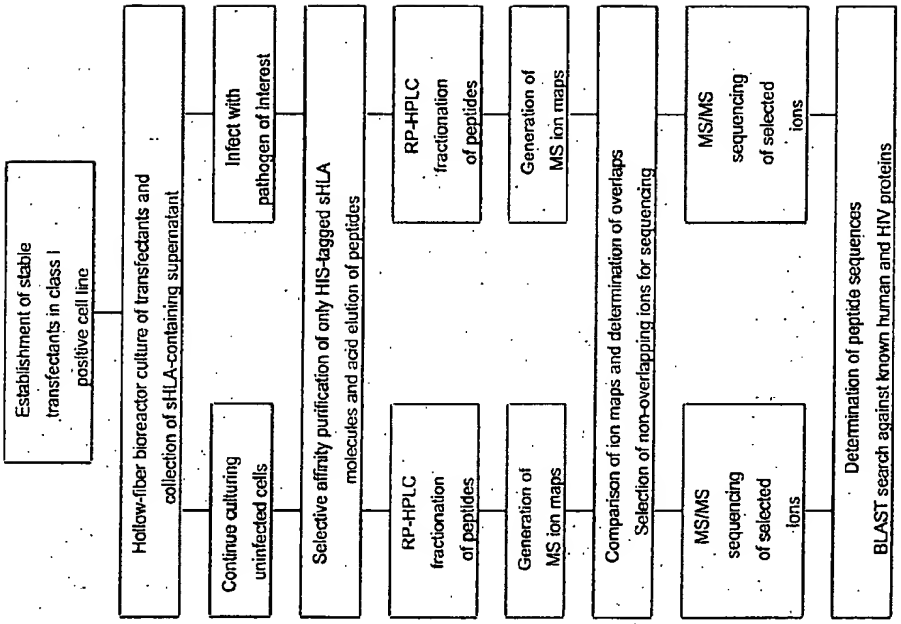

B*1501
B*1501-HIS
B*1501-FLAG
Intensity, cps
m/z, amu
3.03e4 cps
2.70e4 cps
561.1
652.2
765.3
836.4
468.0
383.1
284.1
143.1
86.1
171.0

FIGURE 5


Establishment of stable transfectants in class I positive cell line
Hollow-fiber bioreactor culture of transfectants and collection of sHLA-containing supernatant
Continue culturing uninfected cells
Infect with pathogen of interest
Selective affinity purification of only HIS-tagged sHLA molecules and acid elution of peptides
RP-HPLC fractionation of peptides
RP-HPLC fractionation of peptides
Generation of MS ion maps
Generation of MS ion maps
Comparison of ion maps and determination of overlaps
Selection of non-overlapping ions for sequencing
MS/MS sequencing of selected ions
MS/MS sequencing of selected ions
Determination of peptide sequences
BLAST search against known human and HIV proteins

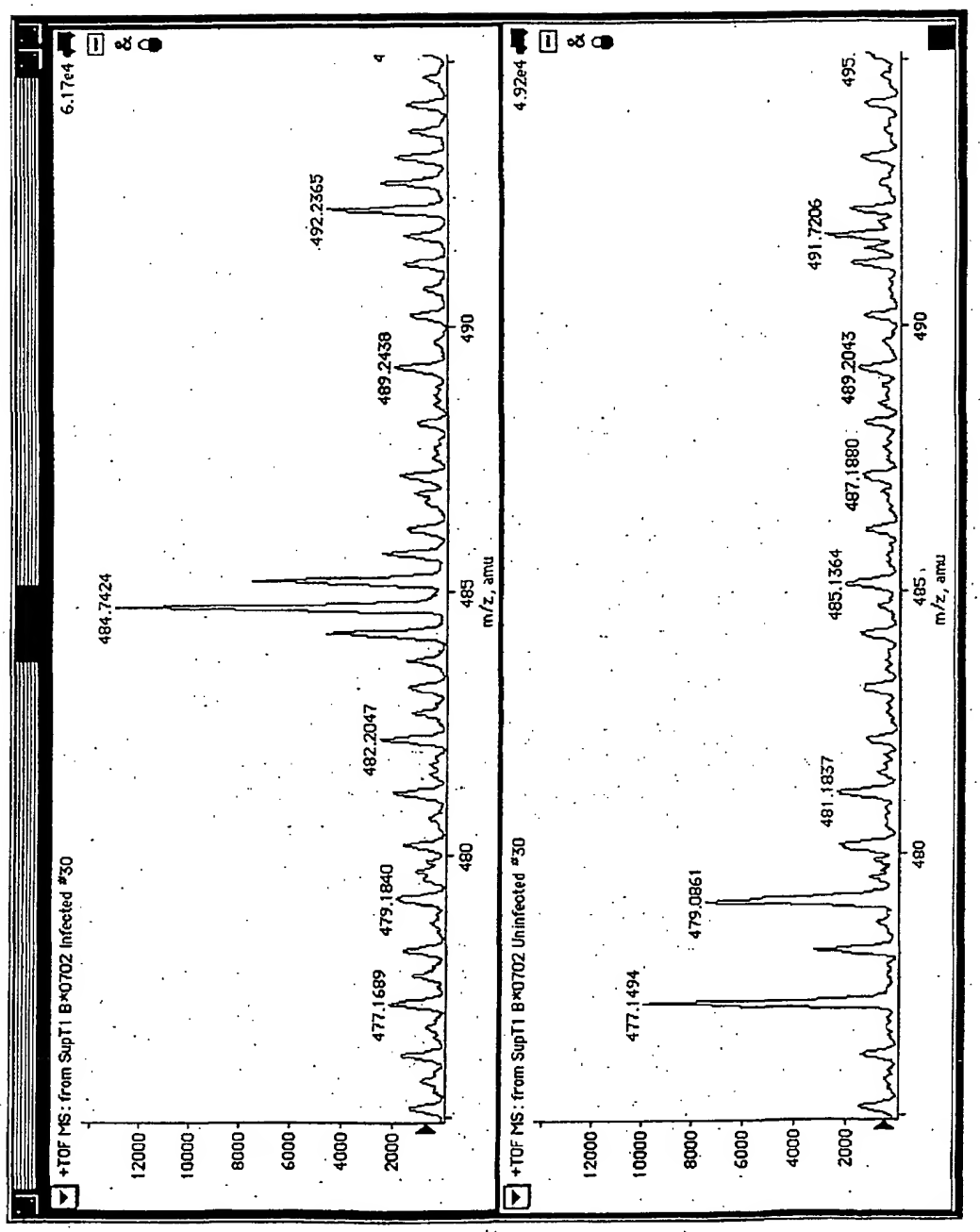

+TOF MS: from SupT1 B*0702 Infected #30
6.17e4
12000
10000
8000
6000
4000
2000
477.1689
479.1840
482.2047
484.7424
489.2438
492.2365
480
485
490
m/z, amu
+TOF MS: from SupT1 B*0702 Uninfected #30
4.92e4
12000
10000
8000
6000
4000
2000
477.1494
479.0861
481.1837
485.1364
487.1880
489.2043
491.7206
495.
480
485
490
m/z, amu

Figure 6

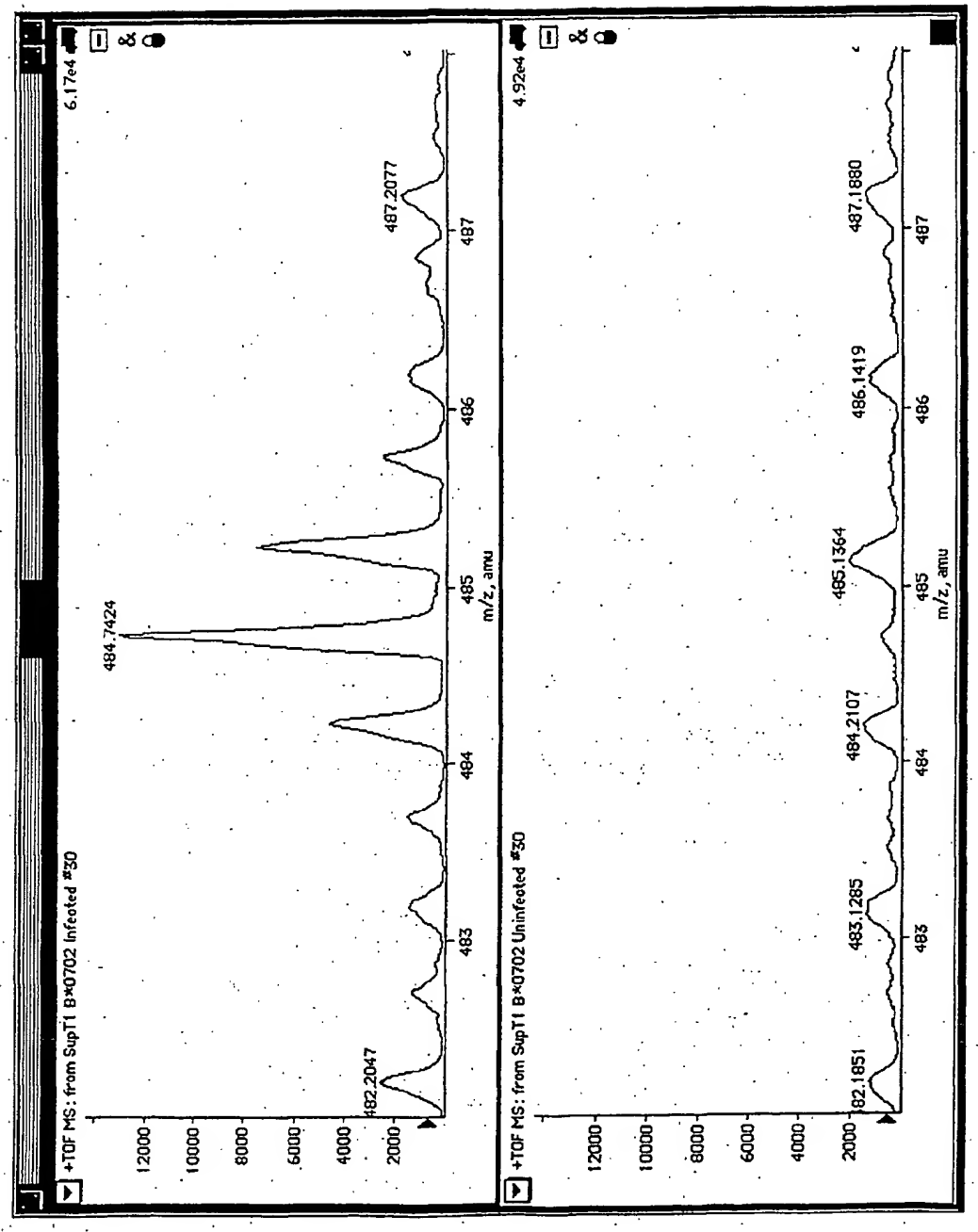

+TOF MS: from Sup11 B*0702 Infected #30
6.17e4
484.7424
487.2077
482.2047
m/z, amu
+TOF MS: from Sup11 B*0702 Uninfected #30
4.92e4
485.1364
486.1419
487.1880
484.2107
483.1285
482.1851
m/z, amu

Figure 7

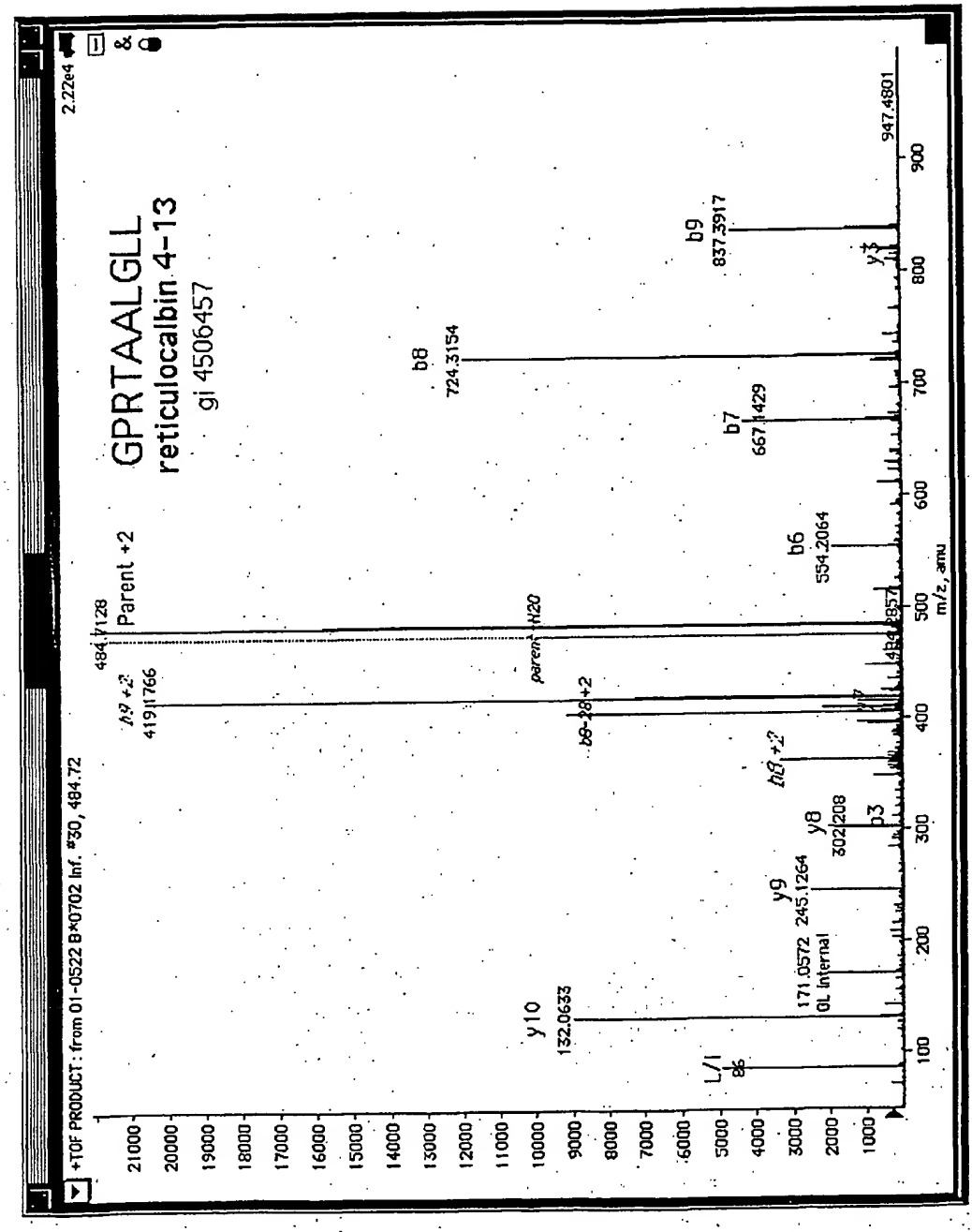

+TOF PRODUCT: from 01-0522 B*0702 Inf. #30, 484.72
2.22e4
GPRTAALGLL
reticulocalbin 4-13
gi 4506457
Parent +2
484.7128
parent-H20
b8
724.3154
b9
837.3917
b7
667.1429
b6
554.2064
y10
132.0633
y9
245.1264
y8
302.1208
b3
171.0572
QL internal
L/I
86
947.4801
m/z, amu

Figure 8.

LIGAND	GPRTAALGLL
Source HLA/Purification	HLA-B*0702 from Sup-T1 subclone B*0702tHIS subclone 2.44 Brandee Run D Purification number 1 (W6/32) Fraction 30 Infected
Protein Source	RETICULOCALBIN 2
GI or accession #	4506457
Ligand start amino acid	4
Ligand length	Decamer
Predicted binding (Parker)	800.000

Protein information from GI

LOCUS NP_002893 317 aa PRI 31-OCT-2000

DEFINITION reticulocalbin 2, EF-hand calcium binding domain; Reticulocalbin 2, EF-hand calcium binding domain (endoplasmic r reticulum calcium-binding protein, 55kD) [Homo sapiens].

COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from 8669.1. Summary: Reticulocalbin 2 is a calcium-binding protein located in the lumen of the ER. The protein contains six conserved regions with similarity to a high affinity Ca(+2)-binding motif, the EF-hand. The RCN2 gene maps to the same region as type 4 Bardet-Biedl syndrome (MIM:600374), suggesting a possible causative role for reticulocalbin 2 in the disorder.

Protein sequence

1 mrlgprtaal gllllcaaaa gagkaeelhy plgerrsdyd reallgvqed vdeyvklghe
61 eqqkrlqaii kkidldsdgf lteselsswi qmsfkhyamq eakqqfveyd knsddtvtwd
121 eyniqmydrv idfdentald daeeesfrkl hlkdkkrfek anqdsgpgls leefiafehp
181 eevdymtefv iqealeehdk ngdgfvslee flgdyrwdpt anedpewilv ekdrfvndyd
241 kdndgrldpq ellpwvvpnn qgiaqeealh lidemdlngd kklseeeile npdlfltsea
301 tdygrqlhdd yfyhdel

FIG 9

Entrez-Pubmed listings:

Carper D, John M, Chen Z, Subramanian S, Wang R, Ma W, Spector A. Gene expression analysis of an H(2)O(2)-resistant lens epithelial cell line. Free Radic Biol Med. 2001 Jul 1;31(1):90-7.

Nimmrich I, Erdmann S, Melchers U, Finke U, Hentsch S, Moyer MP, Hoffmann I, Muller O. Seven genes that are differentially transcribed in colorectal tumor cell lines. Cancer Lett. 2000 Nov 10;160(1):37-43.

Yu LR, Zeng R, Shao XX, Wang N, Xu YH, Xia QC. Identification of differentially expressed proteins between human hepatoma and normal liver cell lines by two-dimensional electrophoresis and liquid chromatography-ion trap mass spectrometry. Electrophoresis. 2000 Aug;21(14):3058-68

Honore B, Vorum H. The CREC family, a novel family of multiple EF-hand, low-affinity Ca(2+)-binding proteins localised to the secretory pathway of mammalian cells. FEBS Lett. 2000 Jan 21;466(1):11-8. Review.

Kent J, Lee M, Schedl A, Boyle S, Fantes J, Powell M, Rushmere N, Abbott C, van Heyningen V, Bickmore WA. The reticulocalbin gene maps to the WAGR region in human and to the Small eye Harwell deletion in mouse. Genomics. 1997 Jun 1;42(2):260-7.

Liu Z, Brattain MG, Appert H. Differential display of reticulocalbin in the highly invasive cell line, MDA-MB-435, versus the poorly invasive cell line, MCF-7. Biochem Biophys Res Commun. 1997 Feb 13;231(2):283-9.

Tachikui H, Navet AF, Ozawa M. Identification of the Ca(2+)-binding domains in reticulocalbin, an endoplasmic reticulum resident Ca(2+)-binding protein with multiple EF-hand motifs. J Biochem (Tokyo). 1997 Jan;121(1):145-9.

Weis K, Griffiths G, Lamond AI. The endoplasmic reticulum calcium-binding protein of 55 kDa is a novel EF-hand protein retained in the endoplasmicreticulum by a carboxyl-terminal His-Asp-Glu-Leu motif. J Biol Chem. 1994 Jul 2;269(29):19142-50.

Chen JJ, Reid CE, Band V, Androphy EJ. Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. Science. 1995 Jul 28;269(5223):529-31.

Summary of Entrez-pubmed entries:

Reticulocalbin is a 55kD protein located in the lumen of the ER and, as the name implies, it binds calcium. It is a member of a large family of similar proteins that are found along the secretory pathway in human cells. No one knows the exact function of this protein, but it is hypothesized that it is a calcium-dependent molecular chaperone. Interestingly, this protein has been described as being upregulated in a variety of cancerous cell lines, including colorectal tumors, metastatic breast cancers, and hepatomas. In several instances, there is a multi-fold increase in its expression in the cancerous lines. In the case of breast cancer cell lines, it is found only in the invasive, metastatic lines and thus has been implicated as a potentiator of metastasis. Another interesting point is that this protein was also identified independently as the E6BP, or E6 binding protein, which binds to the human papilloma virus E6 protein. The fact that this protein is an ER resident protein and that the B7-binding peptide is derived from the leader sequence which should be cleaved in the ER gives further support to the likelihood of presentation of this peptide should reticulocalbin be upregulated in any number of cancerous or virally infected cells.

FIGURE 10

User Parameters and Scoring Information	
method selected to limit number of results	explicit number
number of results requested	20
HLA molecule type selected	B7
length selected for subsequences to be scored	10
echoing mode selected for input sequence	Y
echoing format	numbered lines
length of user's input peptide sequence	317
number of subsequence scores calculated	308
number of top-scoring subsequences reported back in scoring output table	20

Scoring Results			
Rank	Start Position	Subsequence Residue Listing	Score (Estimate of Half Time of Disassociation of a Molecule Containing This Subsequence)
1	4	GPRTaALGLL	800.000
2	244	DGRLdPQELL	60.000
3	19	AAGAgKAEEL	36.000
4	220	TANEdPIWIL	18.000
5	268	ALHLiDEHDL	12.000
6	160	KANQdSGPGL	12.000
7	166	GPGLsLEEFI	8.000
8	287	EILEnPDLFL	6.000
9	6	RTAAIGLLLL	4.000
10	260	NQGIeQEEAL	4.000

Peptide is expected to bind to B7 with high affinity. (Estimated half time of dissociation is 800.00)

Figure 11.

PEPTIDE BINDING PREDICTION USING RAMMENSEE'S SYPEITHI PREDICTION:

Search Report

Return to search conditions
HLA-B*0702 decamers

go to top

HLA-B*0702 decamers

Pos	1 2 3 4 5 6 7 8 9 0	Score
4	G **P** R T A A L G L L	22
166	G **P** G L S L E E F I	18
291	N **P** D L F L T S **E** A	18
248	D **P** Q E L L P W **V** V	17
1	M **R** L G P R T A **A** L	15
6	R **T** A A L G L L **L** L	15
35	R **R** S D Y D R E **A** L	15
199	D **K** N G D G F V **S** L	15
19	A **A** G A G K A E **E** L	14
287	E **I** L E N P D L **F** L	14
5	P **R** T A A L G L **L** L	13
30	Y **P** L G E R R S D Y	13
36	R **S** D Y D R E A **L** L	13
48	Q **E** D V D E Y V **K** L	13
72	K **I** D L D S D G **F** L	13
130	V **I** D F D E N T **A** L	13

Peptide is predicted to bind HLA-B*0702 with a high affinity. Rammensee's prediction scores this peptide 22.

Figure 12.

PREDICTED PROTEASOMAL CLEAVAGE OF COMPLETE RETICULOCALBIN PROTEIN BY PA PRO C:

PAProC predicts the following (109) proteasomal cleavages (made by human proteasome type III) in Name (317 amino acids)

1	MRLGPRTA | ALGLLLLCAAAA
21	GAGK | AE | ELHY | PLGE | RRSDYD |
41	REALLGV | Q | EDVD | EYV | K | L | GHE |
61	E | QQKR | L | Q | AIIKK | I | DLD | SDG | F
81	L | TE | SEL | S | SW | IQM | S | FKH | YA | M | Q |
101	EA | KQQ | FV | EYD | KN | S | D | DTVTW | D
121	EY | NIQMYDRV | I | DF | D | ENTALD |
141	D | AE | EE | S | F | R | K | L | HLK | D | KKR | FE | K
161	A | N | QDSG | PGL | SL | E | EFI | AF | E | H | P |
181	EEVD | YMTEF | V | I | Q | E | ALEE | HDK
201	NGDGFV | SL | E | EFL | GD | Y | RWDPT
221	AN | EDPEWILV | E | KD | RFV | ND | YD
241	KD | NDG | RLDPQELL | PWVVPNN
261	QGIAQ | E | EALHLI | D | EMDLNG | D
281	KK | L | SE | E | EILENPDLFL | T | SE | A
301	T | DYG | RQL | HD | DY | FYHDEL

The proteasome is not predicted to cleave this protein to produce this peptide.

Figure 13.

PREDICTED PROTEASOMAL CLEAVAGE OF COMPLETE RETICULOCALBIN PROTEIN BY NETCHOP 2.0:

Net Chop 2.0 Prediction Results

The predictions for proteasome clevage sites of 1 sequence
Threshold used: 0.5
********** NetChop predictions **********

317 Sequence

```
mrlgprtaalgllllcaaaagagkaeelhyplgerrsdydreallgvqedvdeyvklgheeqqkrlqaiikkidldsdgf
S......S.S.SS.S.S.S..........S..S.......S...S.......S.S....S...S....SSS.......
lteselsswiqmsfkhyamqeakqqfveydknsddtvtwdeyniqmydrvidfdentalddaeeesfrklhlkdkkrfek
..S..S.....S......S......S..S.........S......SS........S.SS.........SS.S..S....S
anqdsgpglsleefiafehpeevdymtefviqealeehdkngdgfvsleeflgdyrwdptanedpewilvekdrfvndyd
...S....S.S..S..............S....SS..........S.S..S...S.S...S........S........S.
kdndgrldpqellpwvvpnnggiaqeealhlidemdlngdkklseeeilenpdlfltseatdygrqlhddyfyhdel
S.S...S....SS.................SS..S.S.....S...........S......S...S......SS.S
```

The proteasome is predicted to cleave after the C-terminal L and may produce a precursor with N-terminal extension of "mrl". (for more information on this form of cleavage see Cascio P, Hilton C, Kisselev AF, Rock KL, Goldberg AL, 2001, 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide, EMBO J, May 15;20(10):2357-66.)

Figure 14.

OTHER PEPTIDES PREDICTED TO BE PRESENTED FROM THIS PROTEIN:

A*0201

Nonamres

1	2	3	4	5	6	7	8	9	Score
R	L	G	P	R	T	A	A	L	27
A	L	L	G	V	Q	E	D	V	26
I	L	V	E	K	D	R	F	V	25
T	A	A	L	G	L	L	L	L	23
R	L	Q	A	I	I	K	K	I	23
I	L	E	N	P	D	L	F	L	23
A	L	G	L	L	L	L	C	A	22
G	L	L	L	L	C	A	A	A	21
L	L	L	L	C	A	A	A	A	21
I	A	F	E	H	P	E	E	V	21
Y	D	R	E	A	L	L	G	V	20
S	L	E	E	F	I	A	F	E	20
L	L	C	A	A	A	A	G	A	19
A	G	A	G	K	A	E	E	L	19
I	A	Q	E	E	A	L	H	L	19
R	T	A	A	L	G	L	L	L	18
K	A	E	E	L	H	Y	P	L	18
K	N	G	D	G	F	V	S	L	18
S	L	E	E	F	L	G	D	Y	18

Decamers

1	2	3	4	5	6	7	8	9	0	Score
A	M	Q	E	A	K	Q	Q	F	V	22
F	I	A	F	E	H	P	E	E	V	22
G	I	A	Q	E	E	A	L	H	L	22
A	L	H	L	I	D	E	M	D	L	22
R	T	A	A	L	G	L	L	L	L	21
A	L	G	L	L	L	L	C	A	A	21
L	L	L	C	A	A	A	A	G	A	21
Y	M	T	E	F	V	I	Q	E	A	21
I	A	Q	E	E	A	L	H	L	I	21
G	L	L	L	L	C	A	A	A	A	20
W	I	L	V	E	K	D	R	F	V	20
K	L	S	E	E	E	I	L	E	N	20
E	I	L	E	N	P	D	L	F	L	20
A	A	L	G	L	L	L	L	C	A	19
K	I	D	L	D	S	D	G	F	L	19
V	I	D	F	D	E	N	T	A	L	19
I	L	E	N	P	D	L	F	L	T	19
A	A	G	A	G	K	A	E	E	L	18
F	L	T	E	S	E	L	S	S	W	18
K	A	N	Q	D	S	G	P	G	L	18
L	I	D	E	M	D	L	N	G	D	18
M	R	L	G	P	R	T	A	A	L	17
L	L	L	L	C	A	A	A	A	G	17

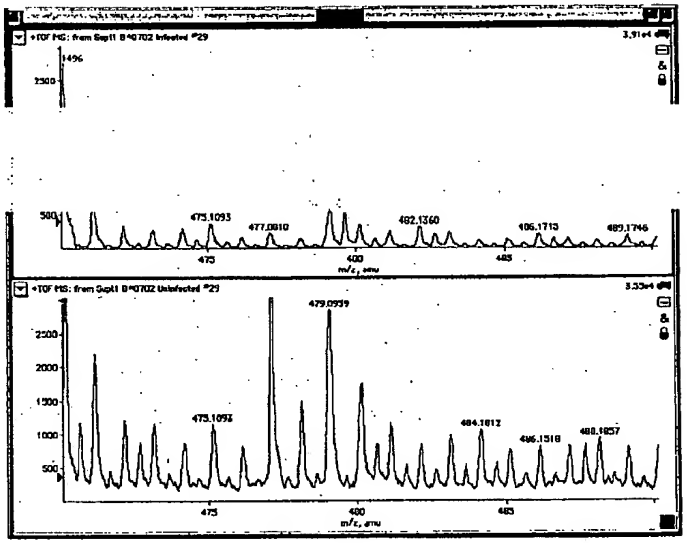
Figure 15. A*0201

A1
Nonamres
1 2 3 4 5 6 7 8 9
Score
S L E E F L G D Y 28
W D E Y N I Q M Y 26
E K D R F V N D Y 26
R L D P Q E L L P 25
A G K A E E L H Y 22
A T D Y G R Q L H 22
E A K Q Q F V E Y 20
F E H P E E V D Y 20
N E D P E W I L V 20
L S E E E I L E N 20
V D E Y V K L G H 19
D Y D R E A L L G 18
W I Q M S F K H Y 18
M T E F V I Q E A 18
I L E N P D L F L 18
Decamers
1 2 3 4 5 6 7 8 9 0
Score
S D D T V T W D E Y 27
T W D E Y N I Q M Y 26
A F E H P E E V D Y 26
A T D Y G R Q L M D 23
G A G K A E E L H Y 21
N E D P E W I L V E 21
R L D P Q E L L P W 21
D L D S D G F L T E 20
S W I Q M S F K H Y 20
V S L E E F L G D Y 20
Y P L G E R R S D Y 19
E S E L S S W I Q M 18
Q E A K Q Q F V E Y 18
A E E L H Y P L G E 17
H E E Q Q K R L Q A 17
N Q D S G P G L S L 17
N G D G F V S L E E 17
A N E D P E W I L V 17
V E K D R F V N D Y 17
D K D N D G R L D P 17
A Q E E A L H L I D 17

Figure 15. A1

View of previous fraction showing 484.72:
Fraction 29

No evidence of the peptide in this fraction.



View of post fraction showing 484.72:
Fraction 31

No evidence of the peptide in this fraction.

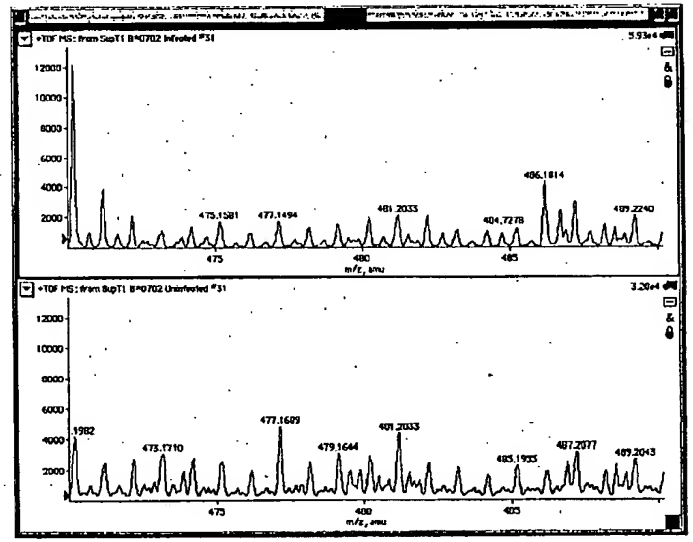


Figure 16

MS/MS of uninfected fraction 30 ion 484.72 under identical collision conditions to prove ion not present:

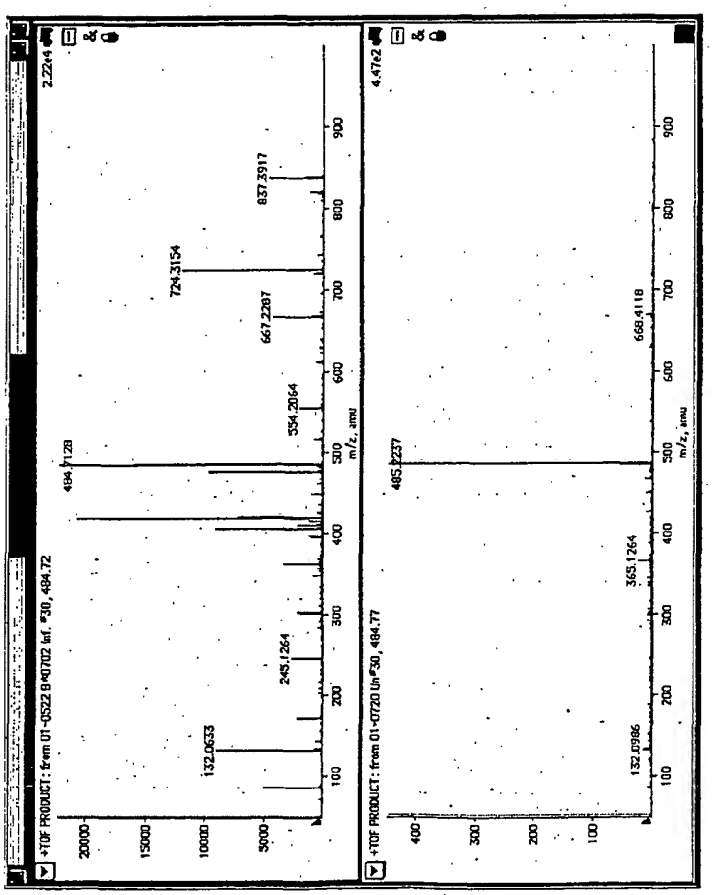


Figure 17.

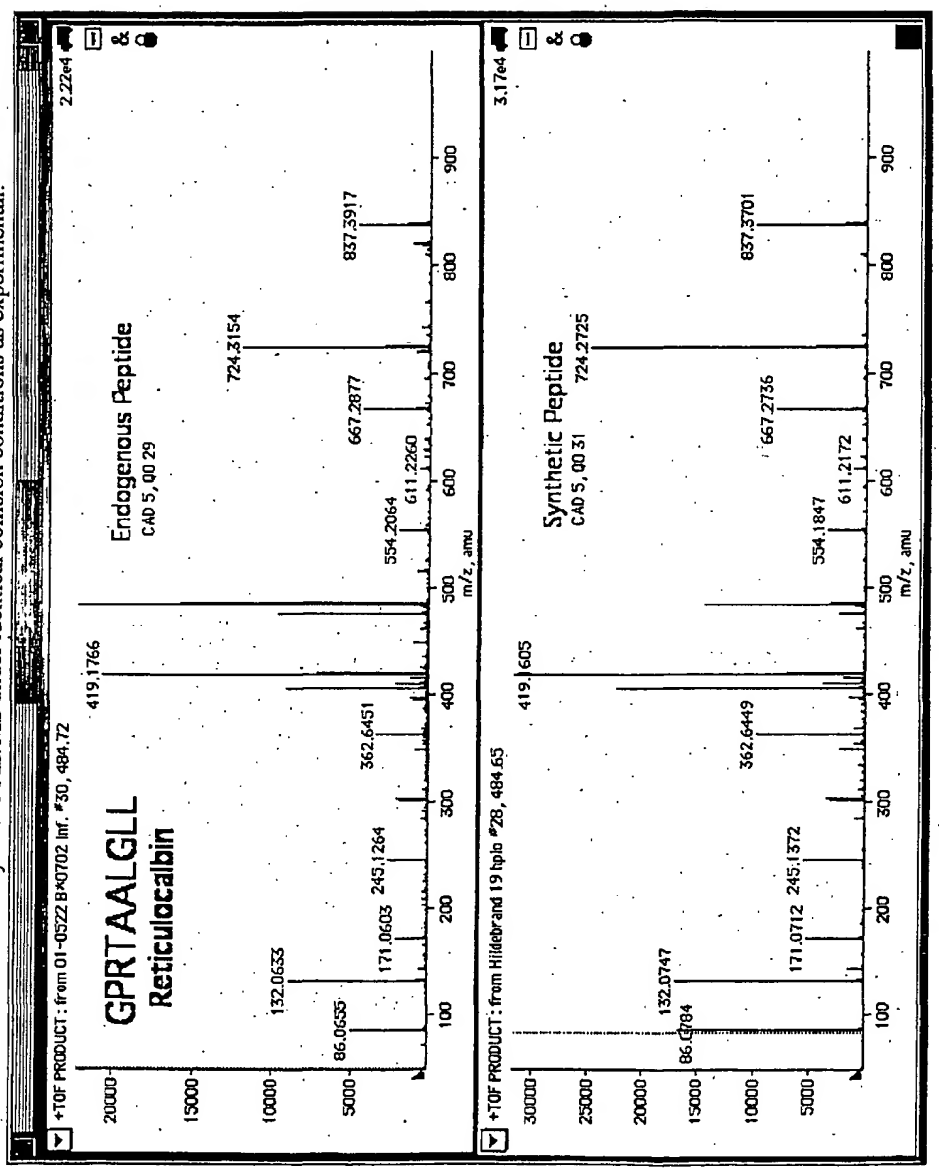

Synthetic MS/MS under identical collision conditions as experimental:
+TOF PRODUCT : from 01-0522 BA0702 Inf. #30, 484.72
GPRTAALGLL
Reticulocalbin
Endogenous Peptide
CAD 5, 00 29
2.22e4
419.1766
724.3154
837.3917
667.2877
554.2064
611.2260
362.6451
245.1264
171.0603
132.0633
86.0655
m/z, amu
+TOF PRODUCT : from Hildebrand 19 hplo #28, 484.65
Synthetic Peptide
CAD 5, 00 31
3.17e4
419.1605
724.2725
837.3701
667.2736
554.1847
611.2172
362.6449
245.1372
171.0712
132.0747
86.0784
m/z, amu

Figure 18.